**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% |

🡪**import** numpy **as** np

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**%matplotlib** inline

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',

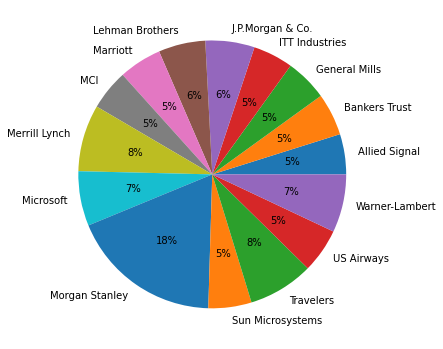
      'Warner-Lambert']

*# Pie Plot*

plt**.**figure(figsize**=**(6,8))

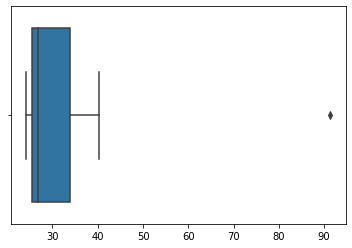
plt**.**pie(x,labels**=**name,autopct**=**'%1.0f%%')

plt**.**show()



**Boxplot for Outliers**

sns.boxplot(data)



**Outliers=91**

**Mean=33.27**

**Variance=287.14**

**SD=16.945**



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.

🡪 inter-quartile range of this dataset IQR=12-5=7 this represents the range whoich contains 50% of the data points.

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

🡪Right 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?

🡪2.5 is not considered as outlier.The box plot will start from 0 and send it to 20 in representation.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

🡪 mode of this dataset lie between 4 and 8

1. Comment on the skewness of the dataset.

🡪Data set is right 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.

🡪 Median in boxplot and Mode in histogram

Histogram provides the frequency distribution so we can see how many times each data point is occurring however boxplot provides the quantile distribution that is 50% data lies

between 5 and 12.

Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an outlier.

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.)

🡪 IF 1 in 200 long-distance telephone calls are getting misdirected.

probability of call misdirecting = 1/200

Probability of call not Misdirecting = 1-1/200 = 199/200

The probability for at least one in five attempted telephone calls reaches the wrong number Number of Calls = 5

n = 5 p = 1/200 q = 199/200

P(x) = at least one in five attempted telephone calls reaches the wrong number P(x) = ⁿCₓ pˣ qⁿ⁻ˣ P(x) = (nCx) (p^x) (q^n-x) # nCr = n! / r! \* (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1 P(1) = **0.0245037**

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?

­🡪 Max. P= 0.3 for P(2000]. So most lkely outcome is 2000$

1. Is the venture likely to be successful? Explain

🡪 Yes, the probability that the venture will make more than 0 or a profit p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit.

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

🡪 Welghted average = x"P("j = 800$. This means the average expected earnings over a long

period of time would be +800$

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

🡪 The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2

= 2800000 – 800^

= 2160000