**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:

**import** numpy **as** np

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

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

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

**%matplotlib** inline

x**=**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])

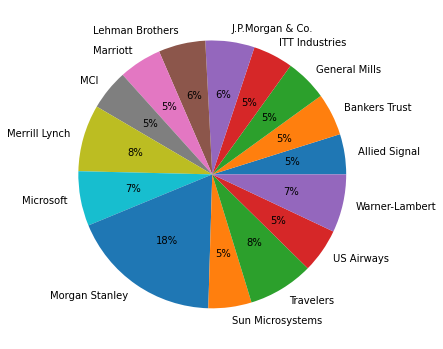
name**=**['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan &Co.','Lehman Brothers','Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways','Warner-Lambert']

*# Pie Plot*

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

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

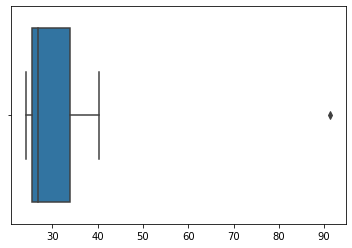
plt**.**show()



*# Box Plot to find outliars*

sns**.**boxplot(x)

<matplotlib.axes.\_subplots.AxesSubplot at 0x2dd3414f730>



*# Mean*

x**.**mean()

33.27133333333333

*# Vairance*

x**.**var()

287.1466123809524

*# Standard Deviation*

x**.**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.

Ans:

Inter-quartile range of this dataset is from 5 to 12 . whisker length is from 0 to19. And 1 outlier.

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

Ans:

Left skewness.

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?

Ans:

It scale the chart 3.

3.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans:

The mode of this dataset lies between 4 to 8.

1. Comment on the skewness of the dataset.

Ans:

Left skewness.

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.

Ans:

We can’t diff mode in box plot but we can do that in Histogram.

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:

One in 200 long-distance telephone calls is misdirected

 probability of call misdirecting  p = 1/200

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

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

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

= 1  -  none of the call reaches the wrong number

= 1  - P(0)

= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1  -  (199/200)⁵

= 0.02475

**probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475**

5.

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:

Most likely monetary outcome of the business venture  is $2000 as it has maximum probability = 0.3

|  |  |  |
| --- | --- | --- |
| E(X) | P(X) | E(X)P(X) |
| -2000 | 0.1 | -200 |
| -1000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
|  |  | 800 |

Expected value =  ∑E(X)P(X)  = 800

long-term average earning of business ventures  = $ 800

venture is  likely to be successful as Expected value is + ve   = $ 800