**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**

xy = pd.DataFrame(index=['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers',

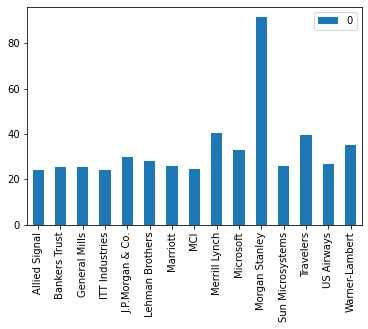
'Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways',

'Warner-Lambert'],data=

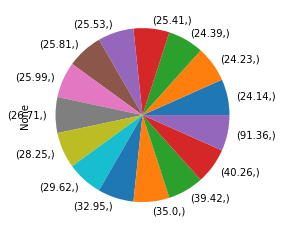
([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00]))

1. **Plot the data**

xy.plot(kind='bar')

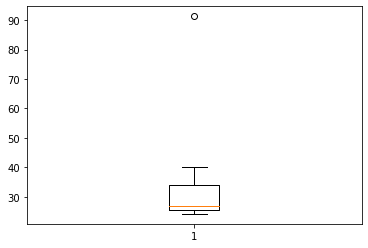


xy.value\_counts().plot(kind='pie')



1. **Find the Outlier**

plt.boxplot(xy)



* **The one outlier at point 92**

xy.mean() **=** **33.271**

xy.var() **= 287.146**

xy.std() = **16.94**





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 – (5 – 12). viscous 0 – 19. 1 outlier. IOR = 7**

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

**ANS – The data are right or positive 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?

**ANS** - **Here 25 is outlier than in new box plot there will be no outlier. It scales the chart.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**ANS – The mode lie between 4 to 8**

1. Comment on the skewness of the dataset.

**ANS – The data is long tail on right side means its positive 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.

**ANS -** **From both histogram and box plot we can find mean, positive skewness, and 25 is outlier. We can’t diff mode in box plot, but we can do that in histogram**

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** - Let us define an event.

E: The call is misdirected then probability of the event E is

Therefore,

P(E)= 1÷200

P(E) = 1-P(E)

= 1- 1÷200

= 199÷200

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

= 1- (199÷200)5

= 1-(0.995)5

= 1 - 0.975

= 0.025

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 -** **The most likely monetary outcome of the business venture: x = 2,000 with the highest probability of 0.3**

1. Is the venture likely to be successful? Explain

**ANS -** The venture is likely to be successful, because

(x = 1,000) + (x = 2,000) + (x = 3,000)

= 0.2+ 0.3 + 0.1 = **0.6**

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

**ANS -** (0.1)(−2,000) + (0.1)(−1,000) + (0.2)(0) + (0.2)(1,000) + 0.3)(1,000) +(0,1)(3,000)

**= 800**

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

**ANS –** Standard Deviation is the good measure of the risk involved in a venture of this kind.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

y =pd.Series([0.1,0.1,0.2,0.2,0.3,0.1])

y.std() = **0.0816**