**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% |

Sol): import pandas as pd

dl=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])

dl.mean() Output: 33.27133333333333

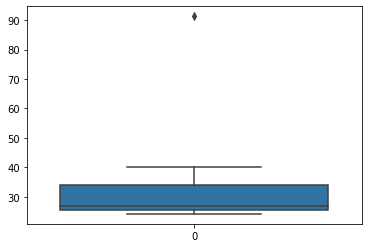
dl.var() Output: 287.1466123809524

dl.std() Output: 16.945400921222028

import seaborn as sns

sns.boxplot(data=dl)

<AxesSubplot:>



Hence outlier in this data is 91.36.



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.

Sol): IQR= upper quartile - lower quartile = 12-5 = 7.

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

Sol): The data is 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?

Sol): Range will differ and the value 2.5 will be considered as the lowest value and it will lie in 1st quartile that is among the 25% of our data. And also its mean, median value might change.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie? Between 4.8 to 8.
2. Comment on the skewness of the dataset. Positively skewed.
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.

Sol: Boxplot: With boxplot we were able to identify minimum, maximum, median and outliers present in it. And 50% of the data lies between 5 to 12. It shows the skewness is positively skewed.

Histogram: With histogram we were able to identify how data is distributed also can identify skewness which is positively skewed. And here we can identify mode of the same dataset by the most occurring value.

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.)

Sol) 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

**So, probability that at least one in five attempted telephone calls reaches the wrong number is 0.02475.**

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?

Sol) It can be seen from the above table that for x = 2000, the value of P(X) is greater. Hence, the most likely monetary outcome of the business venture is x = 2000.

1. Is the venture likely to be successful? Explain

Sol) Yes. Because, P(X>=1000) = 0.6 which is higher than P(X<1000) = 0.4.

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

Sol) Long term average earning of business venture = x.mean()= 500.

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

Sol) Risk = P(X<=0) = 0.2+0.1+0.1 = 0.4.