**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% |

**Solution:**

Here, each parameter is calculated in python. And data visualization technique “Box Plot” is used to identify outliers.

* Plotting of data

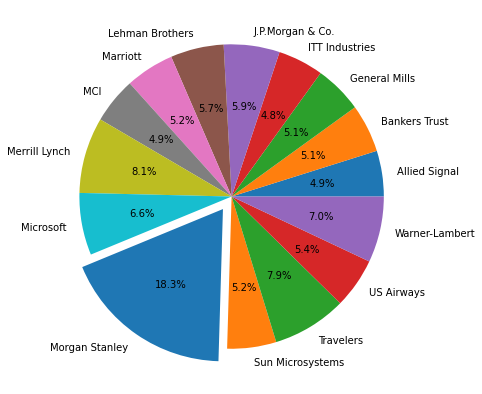


Figure 1.1 Pie plot of given data

* Mean μ= 33.2713
* Standard Deviation σ = 16.945
* Variance σ2 = 287.146
* Box Plot

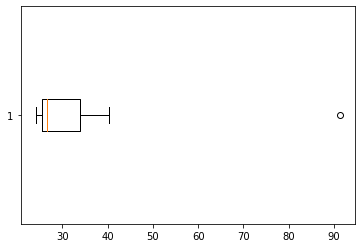
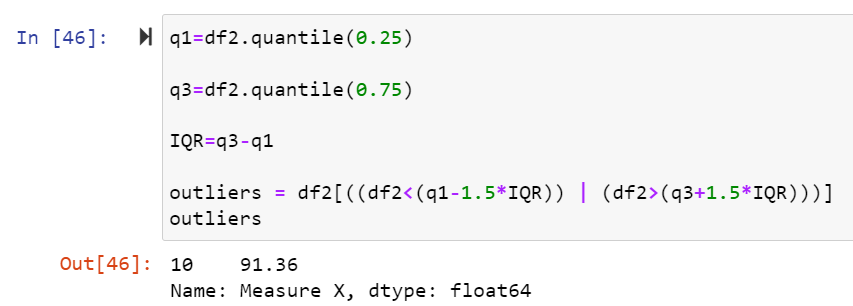


Fig. 1.2 Box Plot

* Outlier- 1



Note: relative python script file is uploaded on github



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.
2. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

**Solution:**

1. Approximately,

First quartile range (Q1) =5

Third quartile range (Q3) =12

Median (second quartile) =7

IQR=Q3-Q1=12-5=7

Comment: IQR is equal to median/second quartile.

1. As observed in above box plot, median is on the left side. So, data distribution is rightly skewed and not normally distributed.
2. There will be no outlier first if data point is 2.5 and also data distribution will be normal as positive skewedness will be reduced.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
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.

**Solution:**

1. The mode of the given dataset lies in between 4 to 8.
2. The above data distribution is rightly skewed, as Mean>Median>Mode.
3. In both plot, right skewedness of data is visible.

In histogram, frequencies of data are displayed that is useful in finding mode.

Whereas, in boxplot, quantile information and median are easily visualized.

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.)

**Solution:**

**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 pˣ qⁿ⁻ˣ where 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?
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

**Solution:**

First, let us find expected value =

|  |  |  |
| --- | --- | --- |
| E(x) | P(x) | E(x)P(x) |
| -2,000 | 0.1 | -200 |
| -1,000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
| SUM | | 800 |

1. What is most likely monetary outcome of the business venture?

The most likely monetary outcome of the business venture is 2000$, as for this the probability is high that is 0.3 among others.

1. Is the venture likely to be successful? Explain

This venture is likely to be successful as expected value of business venture is 800$ that is positive

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

The long-term average earning of business ventures is 800 $

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure
2. 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(x2) –(E(X))2

= 2800000 – 8002

= 2160000