**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% |

**Answer:**

**program:**

import pandas as pd

l={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}

df=pd.DataFrame(l)

print("Mean of the measure",df.mean())

print("standard deviation of the measure",df.std())

print("variance of the measure",df.var())

df.boxplot()

**Output:**

The Mean of the measure 0 33.271333

dtype: float64

standard deviation of the measure 0 16.945401

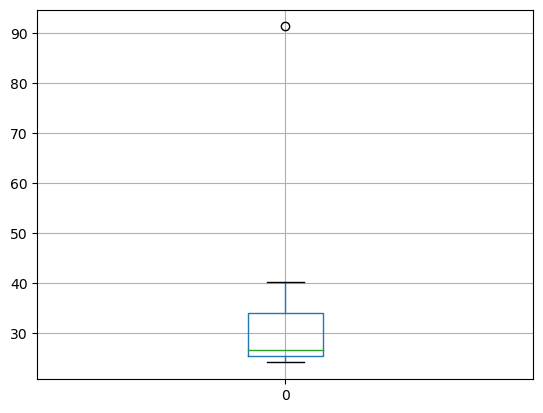
dtype: float64

variance of the measure 0 287.146612

dtype: float64

Out[2]:

<AxesSubplot:>



From the boxplot we can see that there is an outlier. 91.36 point is an outlier in the boxplot.



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.

**Answer:**  From the Box-Plot the inter quartile range is approximately 5 to 12.5. Where in the Q1=5, Q2=6.5, Q3= 12.5 (approximately)

IQR=>q3-q1=>12.5-5=>7.5

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

**Answer:** From the Box plot the data is totally is at left side. So we can conclude that it 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?

**Answer:** If the point of 25 is changed to 2.5. Then the box plot will be changed to without outliers because the point of 25 in the boxplot is an outlier. where it will be changed to 2.5 in the data set we cannot see outliers and we can get values perfectly and we can give 100% perfectly result to the clients if there is no problem with data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Answer:** From the above Histogram, We can say that the data is lies between 1 to 25. Where as Most of the data is lies between 3 to 7.

Hence, the mode is approximately between 3 to 7.

1. Comment on the skewness of the dataset.

**Answer:** From the above histogram the data is positively skewed. Because, most of the data is lies in the left side.

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.

**Answer:** They both are right-skewed and both have outliers the median can be easily visualized in box plot where as in histogram mode is more visible.

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.)

**Answer:**

Probability of a successful call (i.e., not misdirected) is p = 1 - 1/200 = 199/200

Probability of a misdirected call is q = 1/200

Probability of all five calls being successful is:

P(all successful) = (199/200)^5

P(at least one misdirected) = 1 - P(all successful)

= 1 - (199/200)^5 = 0.024 or 2.4%

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?

**Answer:**

From the above table we can say that the outcomes with the highest probability is 2000, with probability of 0.3.

1. Is the venture likely to be successful? Explain

**Answer:**

+ve x values are successful so taking 1000,2000 and 3000.

Probability values are 0.2,0.3,0.1. so P= 0.2+0.3+0.1 = 0.6 > 0.5(p value).so it is

successful.

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

**Answer:**

Expected value=(-2000)\*0.1+(-1000)\*0.1+0\*0.2+1000\*0.2+2000\*0.2+3000\*0.1

=800$

Hence, Long-term average earnings of business venture of this kind is 800$

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

**Answer:** Risk of venture

Var(X)=E(X 2 )-[E(X)] 2

= 2800000-(800) 2

= 2160000 (Quite high)

Std = sqrt(var) = sqrt(2160000)

Std = 1470$

When variate is high then risk is high.