**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% |

1. data=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])

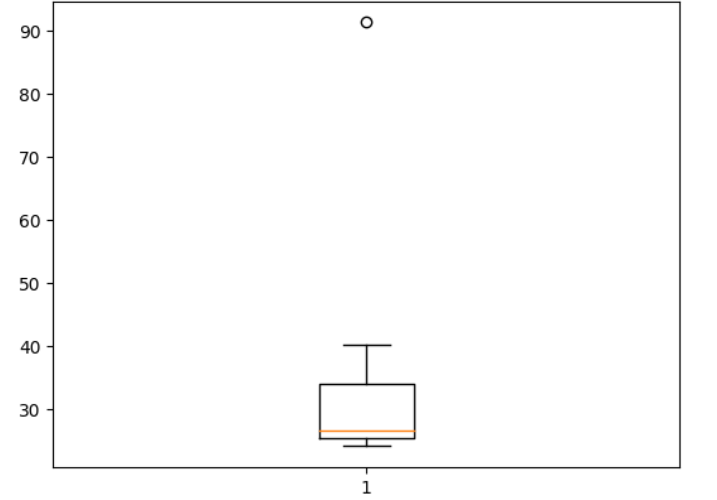
Mean=data.mean()=33.27

Varaiance=data.var()=287.14

Standard Deviation= data.std()=16.94

Plotting the data:

plt.boxplot(data)



Outliers:

From the plotted graph we can see there is only one outlier

Using python code:

import numpy as np

q3,q1=np.percentile(data,[75,25])

iqr=q3-q1

outliers = data[((data<(q1-1.5\*iqr)) | (data>(q3+1.5\*iqr)))]

The output 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.
2. IQR=Upper Quartile-Lower Quartile

=12-5=7

IQR represents the middle 50% of the data i.e. the data which is in the range of 75% to 25%

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

Median is closer to 1st Quartile. So, the data is right 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?

In that case there would be no Outliers on the given dataset because of the outlier the data had positive skewness which will be reduced and the data will be approximately normal distributed



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Mode lie in the range 4-8

1. Comment on the skewness of the dataset.

Dataset is right 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.

A.They both are right-skewed. The main difference is median and outliers can be easily visualized in boxplot whereas it is not possible in histogram but mode can be visualized in histogram which is not possible in boxplot.

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.)
2. IF 1 in 200 long-distance telephone calls are getting misdirected.  
   Probability of call misdirecting = 1/200 =p

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

The probability for at least one in five attempted telephone calls reaches the wrong number Number of Calls 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) (p^x) (q^n-x) # 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. The most likely monetary outcome of the business venture is 2000$. As for 2000$ the probability is 0.3 which is maximum as compared to others.
3. Is the venture likely to be successful? Explain

P(x=-2000)+P(x=-1000)+P(x=0) =0.1+0.1+0.2=0.4

P(x=1000)+ P(x=2000)+ P(x=3000)=0.2+0.3+0.1 =0.6

Here negative and zero returns is less than positive returns (0.4<0.6) So venture is likely to be successful…

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

A. The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 800$

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure
2. Standard deviation is good measure of risk in this kind of venture which is equal to square root of variance

Variance=E(x2) – (E(x))2

E(x2)=4000000\*0.1+1000000\*0.1+0+1000000\*0.2+4000000\*0.3+9000000\*0.1=2800000

E(x)=800

Variance= E(x2) – (E(x))2=2160000

Standard Deviation=sqrt of variance=1469.9=1470

As **Variability is Quite high**  hence **Risk is high**