**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** **pandas** **as** **pd**

**import** **numpy** **as** **np**

In [6]:df\_book1=pd.read\_csv("Book1.csv")

df\_book1

Out[6]:

|  | **Name\_of\_company** | **Measure** |
| --- | --- | --- |
|  |  |  |
| **0** | Allied Signal | 24.23 |
| **1** | Bankers Trust | 25.53 |
| **2** | General Mills | 25.41 |
| **3** | ITT Industries | 24.14 |
| **4** | J.P.Morgan & Co. | 29.62 |
| **5** | Lehman Brothers | 28.25 |
| **6** | Marriott | 25.81 |
| **7** | MCI | 24.39 |
| **8** | Merrill Lynch | 40.26 |
| **9** | Microsoft | 32.95 |
| **10** | Morgan Stanley | 91.36 |
| **11** | Sun Microsystems | 25.99 |
| **12** | Travelers | 39.42 |
| **13** | US Airways | 26.71 |
| **14** | Warner-Lambert | 35.00 |

In [7]:df\_mean=df\_book1.mean()

df\_mean

Out[7]:Measure 33.271333

dtype: float64

In [8]:df\_std=df\_book1.std()

In [9]:df\_std

Out[9]:Measure 16.945401

dtype: float64

In [18]:df\_var=df\_book1.var()

df\_var

Out[18]:Measure 287.146612

dtype: float64

In [31]:**import** **matplotlib.pyplot** **as** **plt**

In [66]:plt.boxplot(df\_book1['Measure '])

Out[66]:{'whiskers': [<matplotlib.lines.Line2D at 0x17e9db45388>,

<matplotlib.lines.Line2D at 0x17e9d6b8188>],

'caps': [<matplotlib.lines.Line2D at 0x17e9d6b8e48>,

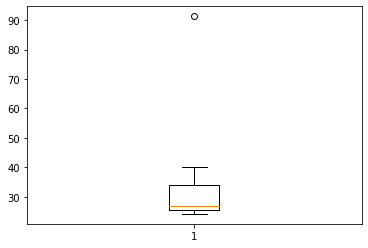
<matplotlib.lines.Line2D at 0x17e9d6b8808>],

'boxes': [<matplotlib.lines.Line2D at 0x17e9d6cf708>],

'medians': [<matplotlib.lines.Line2D at 0x17e9d82b808>],

'fliers': [<matplotlib.lines.Line2D at 0x17e9d82b388>],

'means': []}





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=

IQR=12-5=7

(The interquartile range is the distance between the third and the first quartile, or, in other words, IQR equals Q3 minus Q1)

>50% of the data lies between IQR

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

Ans=The skewness of the dataset is positively skewed

From the above boxplot we can say that the distribution of X is right-skewed

Or 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?

Ans=if we change value 25 with 2.5 then

First quartile(Q1)=2.5

Third quartile(Q3)=15

IQR=Q3-Q1=15-2.5=12.5

Min value=0,max value =20

Here is no any outlier

if it was found that the data point is actually 2.5 instead of 25, the outlier in the boxplot will be removed.

Whether the median shifts or not depends on the size of the data.

It will reduce the right skewness of the data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans= mode=4-8

We need to have actual data to get the exact value of the mode.

The mode can lie between 4 and 10 because there are many values in this range but . this is just an assumption The 2 bars of the same height doesn’t indicate mode

Every time

1. Comment on the skewness of the dataset.

Ans=maximum data is saturated on the left hand side means it is positively 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=we can not differenciate the mode in boxplot but we can do in histogram

25 is outlier in both boxplot and histogram both

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 consider the probability of 1 call misdirected out of 200 as event A.

Probability of occurring of event A= 1/200

P(A)= 1/200

Probability of having at least one successful call will be

1-P(A)= 1-1/200= 199/200= 0.967

As every event is independent of other event the probability will be

1- (0.967)^5

0.02475 = 2% chance.

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=**2000 is the most likely monetary outcome because it has highest probability

1. Is the venture likely to be successful? Explain

Ans=we can say that the venture is success

Because the total probability is 1. Among this ,probability of positive value of

X is 0.2+0.3+0.1=0.6

Which is greater then the probability of negative value ,which is 0.1+0.1=0.2 is likely

be Successful.

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

Ans= long term average is the expected value of the venture =(-2000\*0.1)+

( -1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)

=-200-100+0+200+600+300

=1100-300

=800

Therefore the long-term average earning for these type of ventures would be around 800.

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

Ans=sd=1707.825127659933

