**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 :-

In [4]:

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

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

**import** matplotlib.pyplot **as** plt

**%**matplotlib inline

**import** seaborn **as** sns

In [6]:

x**=**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])

In [9]:

*# variance*

x.var()

Out[9]:

287.1466123809524

In [10]:

*# mean*

x.mean()

Out[10]:

33.27133333333333

In [11]:

*# standard deviation*

x.std()

Out[11]:

16.945400921222028

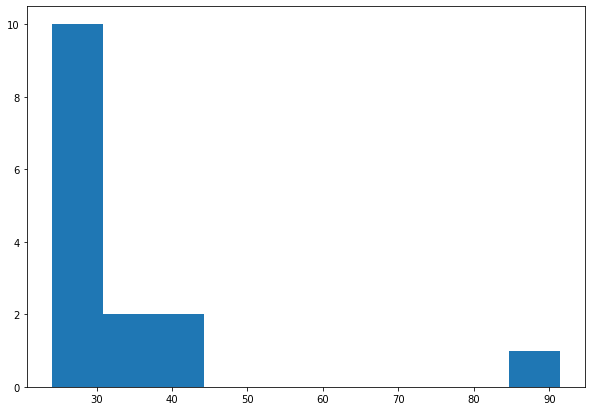
In [13]:

fig**=**plt.figure(figsize**=**(10,7))

plt.hist(x)

Out[13]:

(array([10., 2., 2., 0., 0., 0., 0., 0., 0., 1.]), array([24.14 , 30.862, 37.584, 44.306, 51.028, 57.75 , 64.472, 71.194, 77.916, 84.638, 91.36 ]), <BarContainer object of 10 artists>)



In [16]:

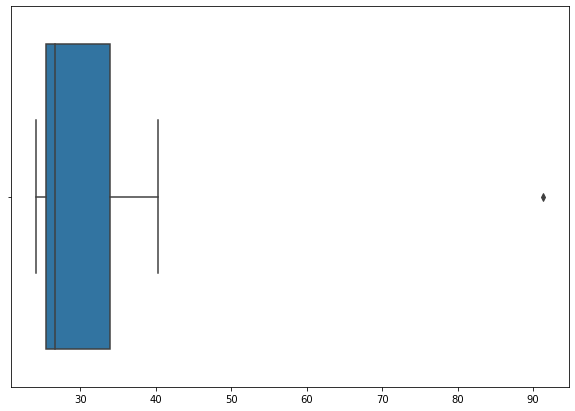
fig**=**plt.figure(figsize**=**(10,7))

sns.boxplot(x)

C:\Users\Rohan Chaure\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation. warnings.warn(

Out[16]:

<AxesSubplot:>



From the above observations,

Morgan Stanley is an outliers of 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.

Ans :- IQR = Q3 – Q1 = 12 – 5 = 7 (approx) ,

IQR represents 50% data.

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

Ans :- Mean is less than median, so the boxplot 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?

Ans :- Outlier in the boxplot will be removed. It will reduce the right skewness of the data.

And with 2.5 (lowest value), IQR = 12 – 2.5 = 9.5



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans :- Between 4-6 and 6-8 (but it cannot be defined exactly without actual data).

1. Comment on the skewness of the dataset.

Ans :- Right skewed (positive).

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 :- From both the graphs we can confirm the outlier at 25 and both indicate ‘+ve’ skewness.

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 :- Probability of occurring 1 misdirected call = p(x) = 1/200

Probability of having at least 1 successful call = 1 – p(x) = 1-1/200 = 0.967

Every event is independent so, 1 – (0.967)^5 = 0.0024 = 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 :- Highest probability = 0.3 with return of $2000.

1. Is the venture likely to be successful? Explain

Ans :- With 60% positive returns (0.2+0.3+0.1=0.6=60%)

And 20% negative returns(0.1+0.1=0.2=20%),

Yes, it is successful.

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

Ans :- Expected Value (EV) = x \* P(x)

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

= 800

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

Ans :- We can use Standard deviation or Variance to measure risk

Variance σ2 = $ 3500000

Standard Deviation σ = $ 1870.829

Compared with standard deviation and average returns it is risky.