**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**:

company\_measures = pd.read\_csv('C:/Users/Ravi Kiran/Basic Statistics Level - 2/Company-Measures.csv')

company\_measures.dtypes

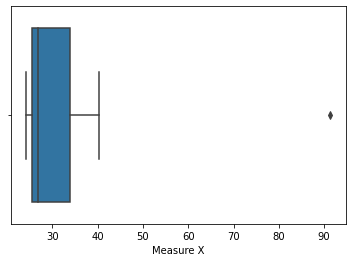
Name of company object

Measure X float64

dtype: object

sns.boxplot(x**=**'Measure X', data**=**company\_measures)

plt.show()



From the box plot, it is observed that **Morgan** **Stanley** **91.36%** is an outlier.

**Mean, Variance, Standard-Deviation**

me\_an**=**company\_measures['Measure X'].mean()

print('Mean =',me\_an)

Mean = 33.27133333333333

variance **=** company\_measures['Measure X'].var()

print('Variance =',variance)

Variance = 287.1466123809524

s\_d **=** company\_measures['Measure X'].std()

print('Standard Deviation =',s\_d)

Standard Deviation = 16.945400921222028



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. IQR = Upper quartile – Lower Quartile

= 12 – 5

= 7

This value implies that Most of the data points lies within the IQR i.e, between the Upper Quartile and Lower Quartile

1. Right – skewed
2. No outliers will be present and Lower quartile might expands so that the IQR will get changed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Solution** : Between the interval 5 - 10

1. Comment on the skewness of the dataset.

**Solution** : 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.

**Solution:**

* Histogram shows frequencies of data points whereas the boxplot doesn’t.
* Boxplot helps to identify outliers easily whereas the histogram doesn’t.

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.)

Probability of call misdirecting p = 1/200

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

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one in five attempted telephone calls reaches the wrong number

= 1 - none of the call reaches the wrong number

= 1 - ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1 - (199/200)⁵

= 0.02475

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?

**2000**

1. Is the venture likely to be successful? Explain

**Solution:**

Positive Returns Probability = 0.2+0.3+0.1 = 0.6 (60%)

Negative Returns Probability = 0.1+0.1 = 0.2 (20%)

More Probability to get positive returns. Hence, the venture is likely to be successful.

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

The long average earning of this business venture will be equal to expected value of x and 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

**Variance**

1

data['x'].var()

Out[46]:

3500000.0

In [47]:

1

data['x'].std()

Out[47]:

1870.8286933869706