Basic Statistics Level-2

Saikrishna Budi

Linkedin : http://linkedin.com/in/sai-krishna-budi-3b0341209

Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

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%

Sol:

Firstly, we will import the required libraries for the above data set.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

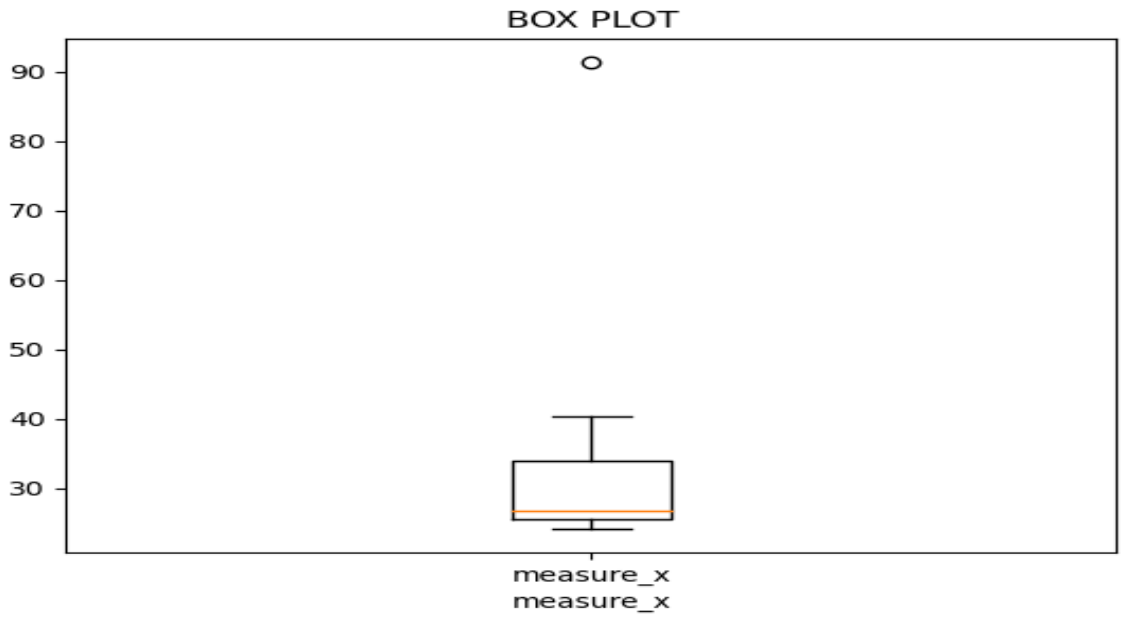
Given data

```
companies = ["Allied Signal", "Bankers Trust", "General Mills", "ITT Industries", "J.P.Morgan &
Co.","Lehman Brothers","Marriott", "MCI", "Merrill Lynch", "Microsoft","Morgan Stanley", "Sun
Microsystems", "Travelers", "USAirways", "Warner-Lambert"]
```

Percentage values (convert to numerical)

```
measure_x = [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]
plt.boxplot(measure_x)
plt.xticks([1],["measure_x"])
plt.xlabel("measure_x")
plt.title("BOX PLOT")
```

plt.show()




```
np.mean(measure_x)
np.std(measure_x)
np.var(measure_x)
```

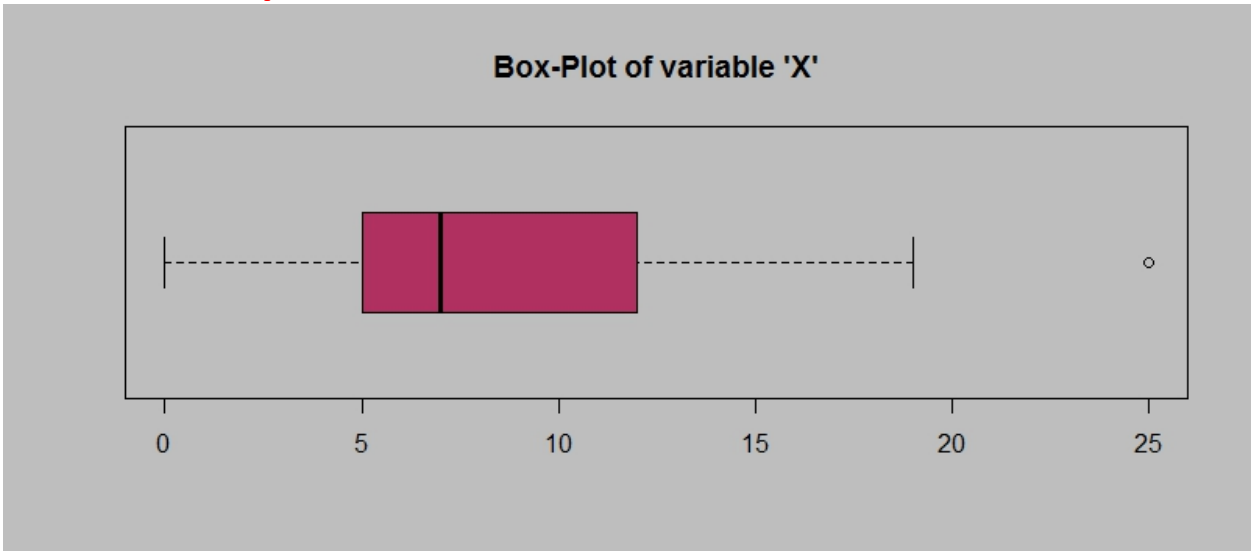
Outliers= Morgan Stanley (91.36)
Mean μ= 33.27133333333333
Standard Deviation σ=16.370812590976932
Variance σ^2=268.00350488888887

2.

Basic Statistics Level-2

Saikrishna Budi

Linkedin : http://linkedin.com/in/sai-krishna-budi-3b0341209



Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

(ii) What can we say about the skewness of this dataset?

(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

Sol:

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

- Here the outlier is 25
  Median=7
  1st quartile =5
  2nd quartile =12
  Inter-Quartile Range (IQR)=Q2-Q1=12-5= 7
  IQR=7

(ii) What can we say about the skewness of this dataset?

- Right Skewed which is positive skewed data.

(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

- There would be no outliers in the data but it could affect the mean and median of the data.

Basic Statistics Level-2

Saikrishna Budi
Linkedin : http://linkedin.com/in/sai-krishna-budi-3b0341209

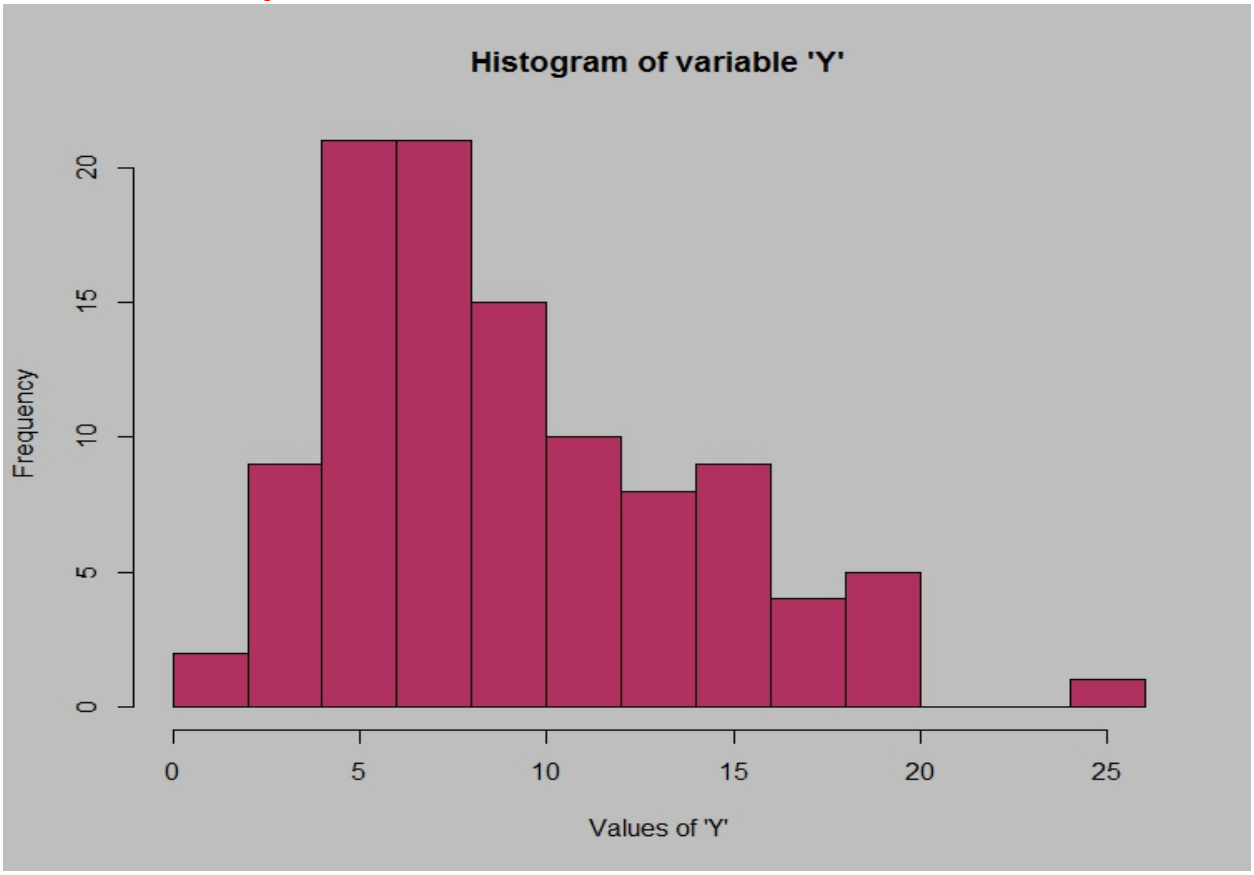
3.

Questions referred to from *Aczel A., Sounderpandian J., Complete Business Statistics (7ed.)*

Basic Statistics Level-2

Saikrishna Budi

Linkedin : http://linkedin.com/in/sai-krishna-budi-3b0341209



Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?

(ii) Comment on the skewness of the dataset.

(iii) 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.

Sol:

(i) Where would the mode of this dataset lie?

- Between 5&8 the mode of dataset lies because most frequent data occur in between it.

(ii) Comment on the skewness of the dataset.

- It is positively skewed data.

(iii) 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.

- By comparing them it is very clear that the data would be positively skewed and it helps in finding the mean, mode of the data.

4. 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.)

Sol:

- To find the chance of a call being misdirected, we divide misdirected calls by total calls, giving a probability of 1/200.
- The probability of a call not being misdirected is 199/200.
- Using the complement rule, we raise this probability to the power of 5 to find the chance that none of five calls are misdirected, which is 0.831.
- Subtracting this from 1 gives the probability of at least one call being misdirected, which is about 0.169.

5. 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

(i) What is the most likely monetary outcome of the business venture?
(ii) Is the venture likely to be successful? Explain
(iii) What is the long-term average earning of business ventures of this kind? Explain
(iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

Sol:

(i) What is the most likely monetary outcome of the business venture?

- The most likely monetary outcome of the business venture is $2,000, as this is the value of x with the highest probability (0.3).

(ii) Is the venture likely to be successful? Explain

- It is not clear whether the venture is likely to be successful based on the given probability distribution. In order to determine whether the venture is likely to be successful, we would need to know the definition of success and whether the expected return is sufficient to meet that definition.

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

- The long-term average earning of business ventures of this kind can be calculated by multiplying each possible outcome by its probability and summing the results.
- Using the given probability distribution, the expected value is (-$2,000 * 0.1) + (-$1,000 * 0.1) + ($0 * 0.2) + ($1,000 * 0.2) + ($2,000 * 0.3) + ($3,000 * 0.1) = $800.
- Therefore, the long-term average earning of business ventures of this kind is $800.

(iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

- The Risk involved in the venture can be calculated by standard deviation and variance by calculating the income (**X*P(X)**.
- Standard Deviation Value is = 294.3
- Variance value is =86612.49