Topics: Descriptive Statistics and Probability

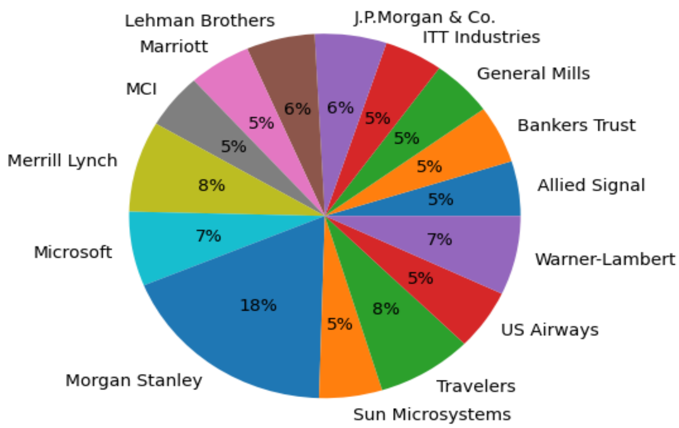
1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

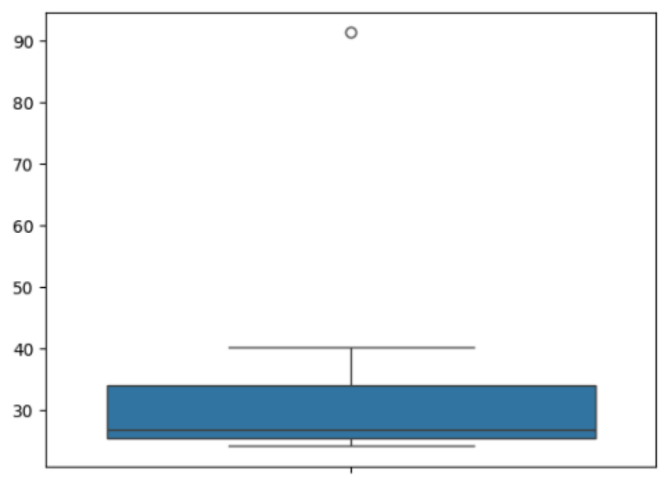
Name of company	**Measure X**
Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%

Ans:

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

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])

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

# Pie Plot
plt.figure(figsize=(5,6))
plt.pie(x,labels=name,autopct='%1.0f%%')
plt.show()
```

Pie chart





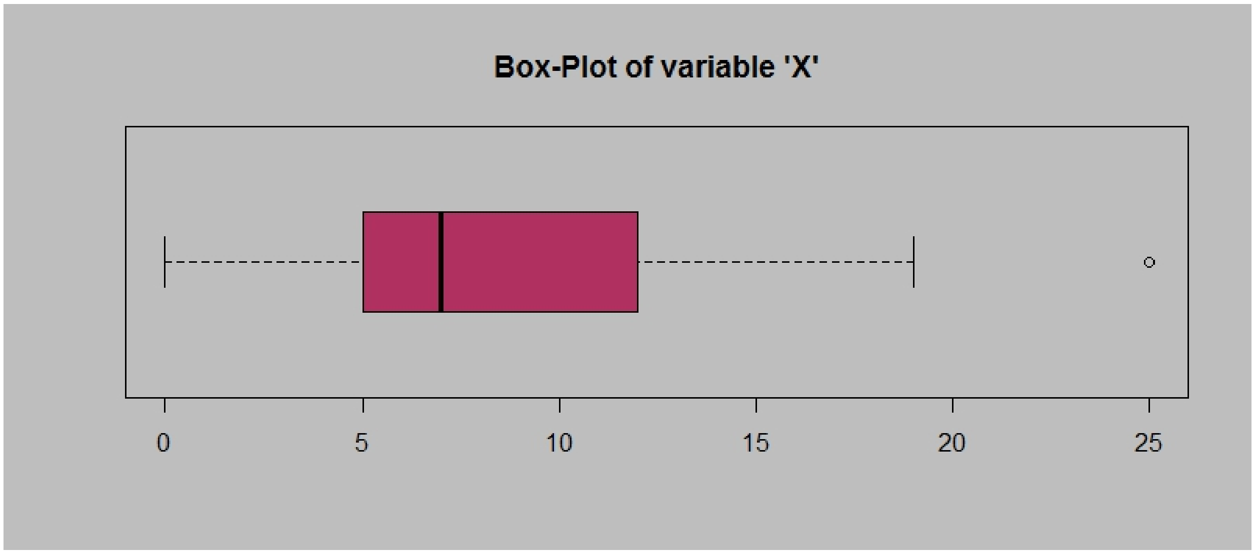
Mean, μ= 33.271

Standard deviation, σ^2 = 16.945

Variance , σ=287.146

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

2.



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.

Ans: The Inter-quartile range of a dataset is the difference between upper quartile and lower Quartile.

Here,

First quartile range Q1 = 5

Second quartile range Q2 = 7 (median)

Third quartile range Q3 = 12

So, the inter-quartile range, IQR = Q3-Q1 = 12-5 = 7

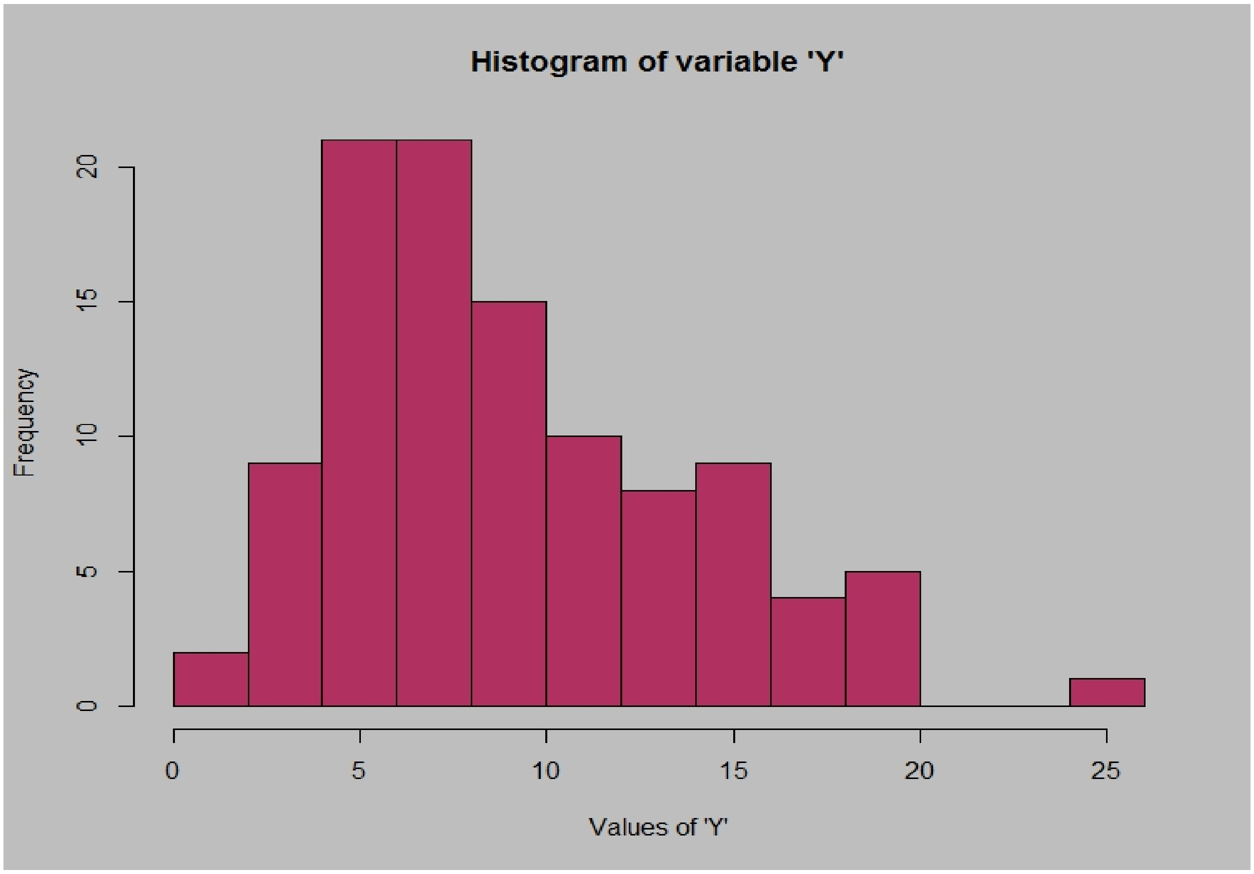
(ii) What can we say about the skewness of this dataset?

Ans: Right-skewed median is towards the left side. It is not normally distributed.

(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?

Ans: In this case there would be no outliers. because of the outliers the data had positive Skewness it will reduce and the data will normally distributed.

3.



Answer the following three questions based on the histogram above.

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

Ans: The mode of this dataset approximately lies between 4 to 8

(ii) Comment on the skewness of the dataset.

Ans: The dataset is positively skewed or right skewed

(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.

Ans: These graphs complement each other. both are positively skewed and both have Outliers. The median can be easily visualized in box plot. and in histogram mode is more Visible.

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.)

Ans:
If 1 in 200 long-distance telephone calls are getting misdirected.

probability of call misdirecting = 1/200

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

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

n = 5

p = 1/200

q = 199/200

P(x) = at least one in five attempted telephone calls reaches the wrong number

$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.0245037

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?

Ans: The most likely monetary outcome of the business venture is 2000$.The probability of 2000$ is 0.3, which is the maximum compared to others.

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

Ans: yes, it's likely to be successful.
p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8
we can see that there is a 80% chance for this venture to be make a profit.

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

Ans: Long term average is Expected value

Expected value = sum(X * P(X)) = (-2000)*0.1+(1000)*0.1+0*0.2+1000*0.2+2000*0.3+3000*0.1

= 800 $

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

Ans: From the above table we can calculate the average return like

Average return = [-2000*0.1] + [1000*0.1] + [0*0.2] + [1000*0.2]+[2000*0.3]+[3000*0.1]

= $800

Standard Deviation = [((-2000-800)^2*0.1) + ((-1000-800)^0.1)+((0-800)^2*0.2((1000-800))^2*0.2) + ((2000-800)^2*0.2) + ((3000-800)^2*0.1)] = $1870.83

As, higher the Standard deviation the greater will be the risk.