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

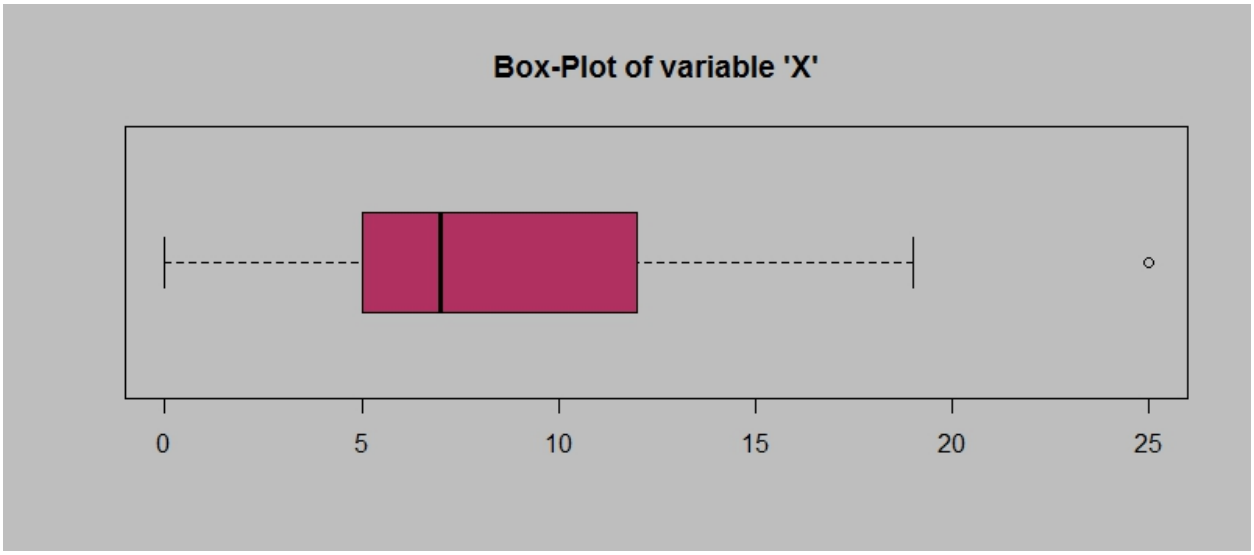
Name of company	Measure X
Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%
Warner-Lambert	35.00%

Ans:

The outliers in the above mentioned table is Morgan Stanley(91.36%)

Mean (μ)=33.27133
Variance(σ^2) = 287.1466
Standard Deviation(σ) = 16.9454

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.

(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?

Ans:

i. It represents the difference between the upperbond and lowerbond values in the shaded region only.

Q1=5

Q3=12

IQR=Q3-Q1

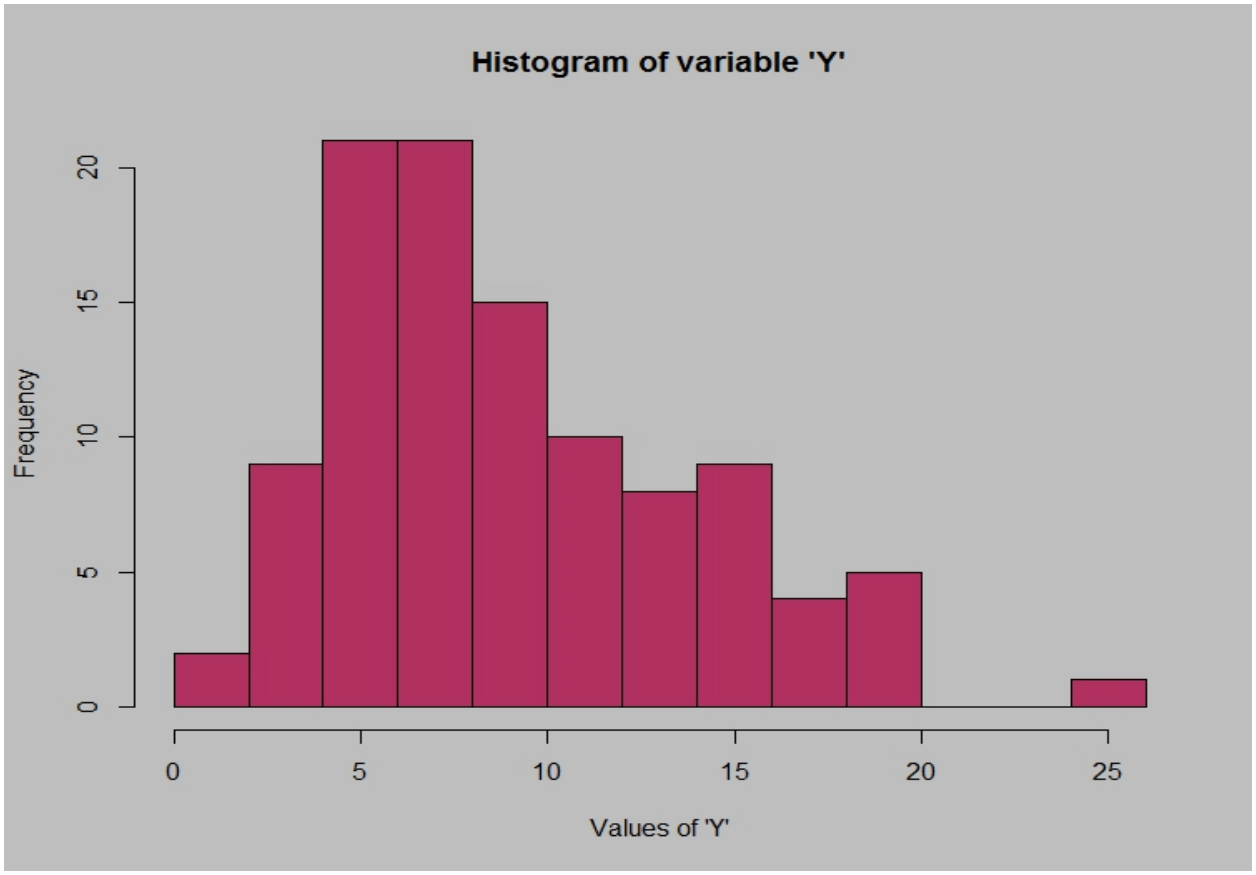
= 12-5

= 7

ii. The above boxplot is representing the Right Skewness or a Positive Skewness of the data

iii. As we can see, the original value (2.5) is less than the existing value (25), as we can also observe that the boxplot is in a positive skewness; hence, even after changing the existing value with the original value, the boxplot is not going to have more effect compared to the present boxplot diagram.

3.



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.

Ans:

i. The mode for the above histogram lies in the range of 4-8.

ii. The skewness of the dataset is going to be positive skew as the long tail points are located on right side.

iii. Both the histogram (question-3) and box-plot (question-2) are positive skewness, these graph offers a complete picture of the dataset. You can visually assess the skewness or symmetry of the data using the histogram,

while the box plot provides a summary of the central location, variability, and potential extreme values.

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:

Given that

The probability of a single call being misdirected (P) = 1/200

The probability of a single call being not misdirected (Q) =1-P

=1-(1/200)

=199/200

Now,

The probability that non of one in five attempted telephone calss reaches the wrong number $=(Q)^5$

$= (199/200)^5$

The probability that one in five attempted telephone calls reached the wrong number

$=1-(Q)^5$

$=1-(199/200)^5$

$=0.02475$

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.

Ans:

i. As we can see the given data that contains the probability distribution for the ventures that are nearest to $1,000, in the entire table we can observe that there is a most likely monetary outcome at the x=2000 with the probability of P(x)=0.3

ii. The venture is running at a successful stage as we can see that the value of x of 1000,2000,3000 having a sum of the probability of 0.6(0.2+0.3+0.1).

iii. As we can see that the given data was having more positive returns with a high probability compared to the negative values and probability, lets calculate the values

⇨ (-2000*0.1) +(-1000*0.1)+(0*0.2)+(1000*0.2)+(2000*0.3)+(3000*0.1)
⇨ -200-100+0+200+600+300
⇨ -300+1100
⇨ 800

iv. A good measure of the risk involved in a venture of this kind is the standard deviation

Let's calculate the standard deviation step by step:

Calculate the squared difference between each value and the mean:

(-2,000 - 700)^2 = 2,724,100

(-1,700 - 700)^2 = 1,000,000

(-700 - 700)^2 = 784,000

(300 - 700)^2 = 160,000

(1,300 - 700)^2 = 160,000

(2,300 - 700)^2 = 1,600,000

Multiply each squared difference by its respective probability:

(2,724,100 * 0.1) = 272,410

(1,000,000 * 0.1) = 100,000

(784,000 * 0.2) = 156,800

(160,000 * 0.2) = 32,000

(1,600,000 * 0.3) = 480,000

(1,600,000 * 0.1) = 160,000

Sum up these values:

272,410 + 100,000 + 156,800 + 32,000 + 480,000 + 160,000 = 1,201,210

Take the square root of the result:

sqrt(1,201,210) ≈ 1,097.32

So, the standard deviation is approximately 1,097.32 for this dataset.