Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ2

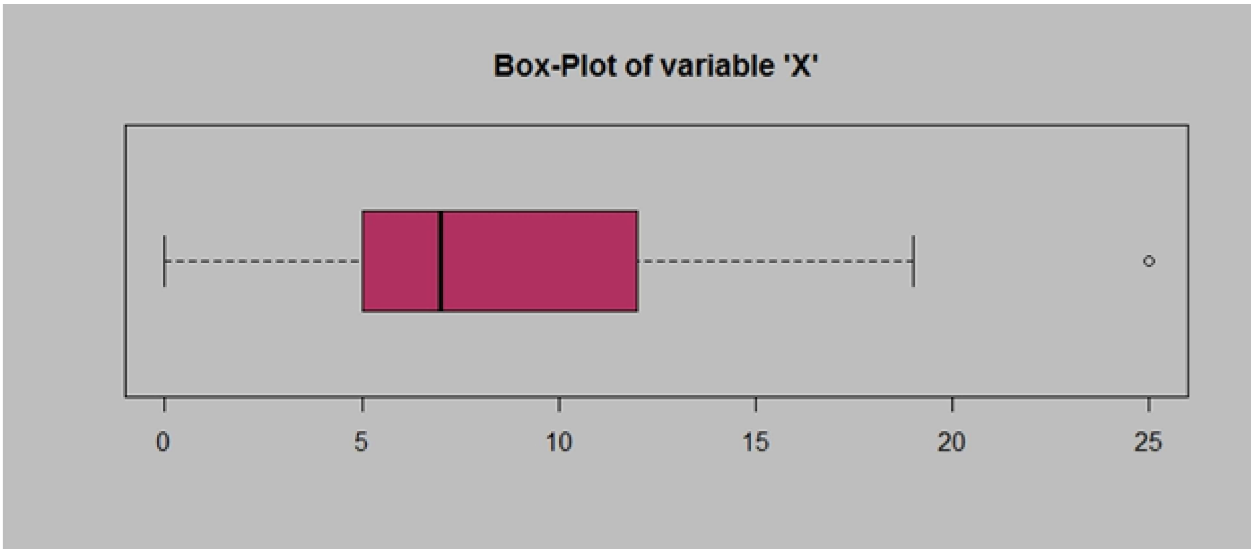
e of company	Measure X
l Signal	24.23%
ers Trust	25.53%
eral Mills	25.41%
ıdustries	24.14%
lorgan & Co.	29.62%
ıan Brothers	28.25%
iott	25.81%
	24.39%
ill Lynch	40.26%
ɔsoft	32.95%
gan Stanley	91.36%
Microsystems	25.99%
elers	39.42%
rways	26.71%
ler-Lambert	35.00%

Q1_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file Set+1_Descriptive stats_Q1.ipynb contains the solution to the above question.

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?

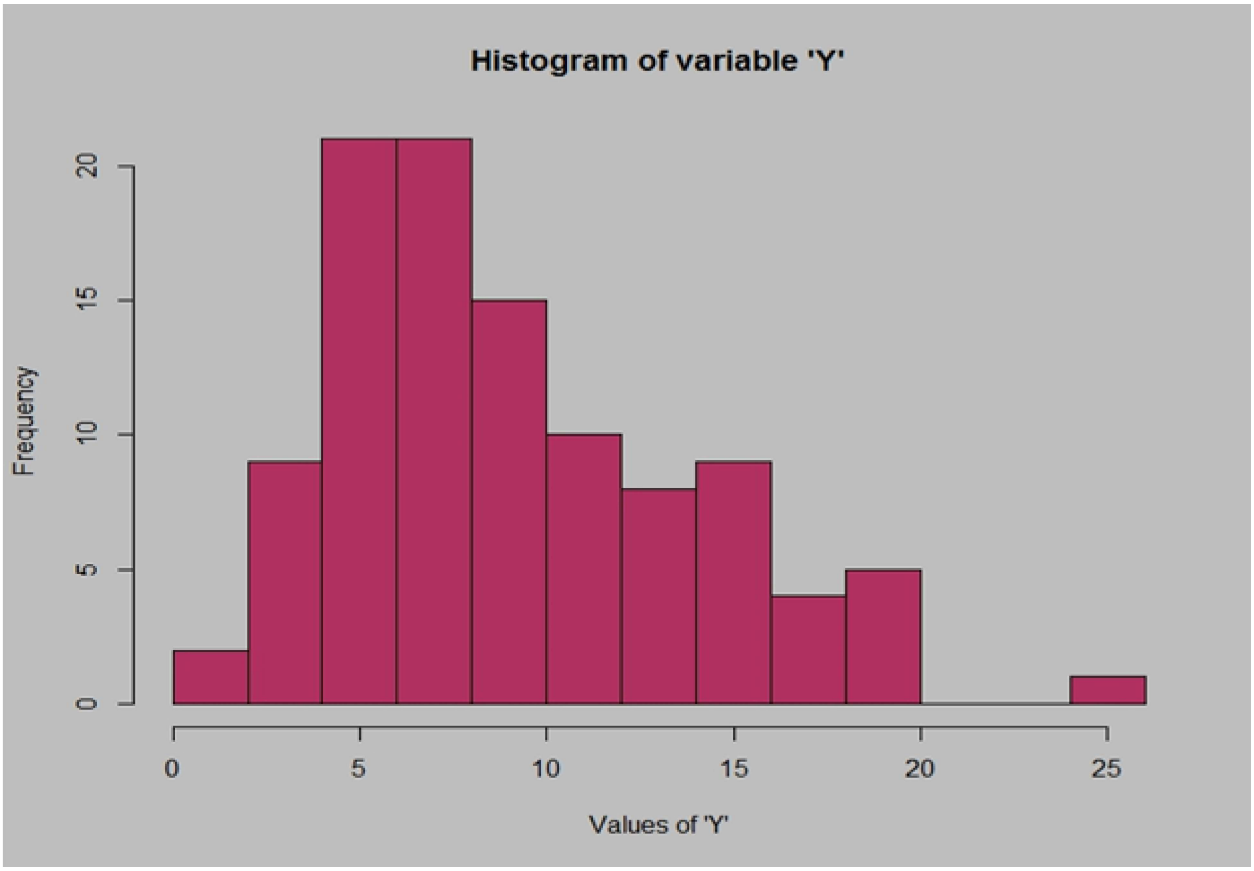
(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

2(1) Solution:
Q1=5, Q3=12 IQR=Q3-Q1=12-5=7
This value implies the median of the data is same as the IQR.

2(2) Solution:
As the median of the data divides the box into two whiskers i.e left and right whiskers
As median is closer to Q1 indicating that the data is more spread out on the right side of the median showing that its right skewed and positively skewed.

2(3) Solution:
There would have been no effect on the box plot if the value 25 was 2.5 it would be still considered as an outlier from the data. because the box plot starts from 5.

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.

3(1) Solution:

The mode of this data set will be 5/6 because mode is known as the most repeated value in the whole data.

3(2) Solution:

As the graph itself is leaned with right side of the data hence its positive skew or right skewed where mode<median<mean hence the skewness>0.

3(3) Solution:

Both the boxplot and the histogram represent that the data is not symmetrical and positively skewed and the values of the dataset are below 25 and starts from zero and there is one value at 25(outlier).

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.)

Q4 Solution:

Handwritten Solution in the next page i.e pageno.5

Q_4 Solution:

Given, one in 200 calls is misdirected.

∴ probability of misdirecting, $p = 1/200$

∴ probability of not misdirecting $q = 1-p$

$= 1 - 1/200 = \frac{199}{200}$

Probability that atleast one call in five attempts reaches the wrong number, $P(X = r) = ?$

Therefore, Binomial probability formula

no. of calls $(n) = 5$

$P(r) = {}^{n}C_r p^r q^{n-r}$,

Here $n = 5$

$p = 1/200$, $q = \frac{199}{200}$

$\therefore P(m) = 1 - P(0)$

Atleast one in five reaches wrong number.

= 1 - none of the calls reaches wrong number.

$= 1 - P(0)$

$= 1 - {}^{5}C_0 p^0 q^{5-0}$ $= 1 - 0 (1/200)^0 \left(\frac{199}{200}\right)^5$

$= 0.02475$

$\therefore P(r) = 0.02475$

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

Q5 Solution:

Handwritten Solution in the next page i.e pageno.7

Q5 Solution :

Probability Distribution table

x	P(x)	E(x) = x.P(x)	E(x²) = x².P(x)
-2000	0.1	-200	400000
-1000	0.1	-100	100000
0	0.2	0	0
1000	0.2	200	200000
2000	0.3	600	1200000
3000	0.1	300	900000
		μ = 800	2800000

(i) the Most likely monetary outcome of the business Venture is the one with the Highest probability - 2000 $

i.e P(0.3) = 2000

(ii) as the probabilities of Earnings is more, i.e P(0.2) + P(0.3) + P(0.1) = P(0.6) ie P(0.6) > P(0.2). Hence the business is profitable

Profitable buisness represent Profit successful Buisness. Loss

(iii) Long Term average Earning is Average, μ = 800.

μ is E(x) = Σx.P(x).

(iii) Standard deviation and Variance, these can calculate Risks.

Var(x) = E(x²) - [E(x)]²

= 2800000 - (800)² = 2160000 .... (High variance)

Stdev = √Var = 1470 $ (if loss incurs then 1470 $ will be lost)