

## Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out  $\mu, \sigma, \sigma^2$

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.

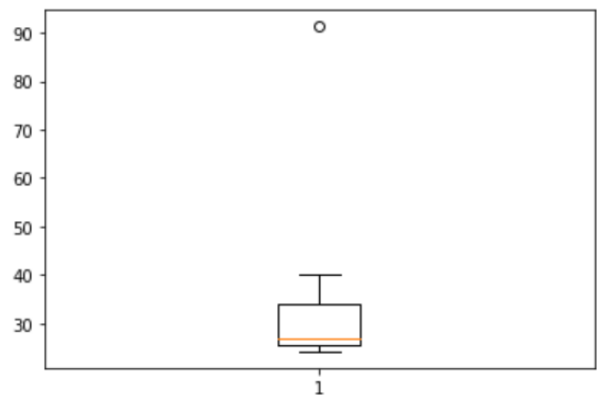
Morgan Stanley with 91.36% is the outlier

For the given data,

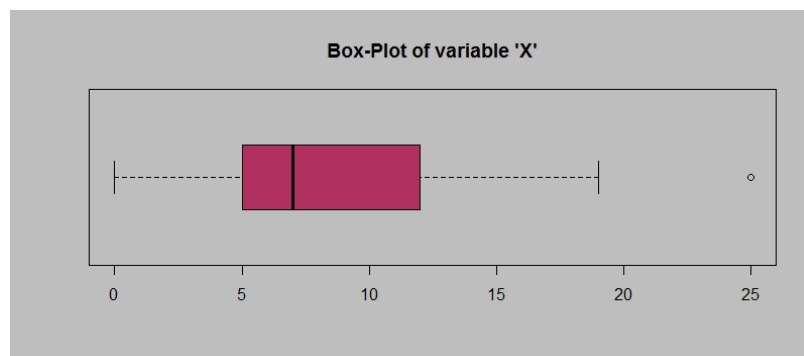
$$\underline{\underline{\mu = 33.271\%}}$$

$$\underline{\underline{\sigma^2 = 287.145}}$$

$$\underline{\underline{\sigma = 16.945\%}}$$



2.



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

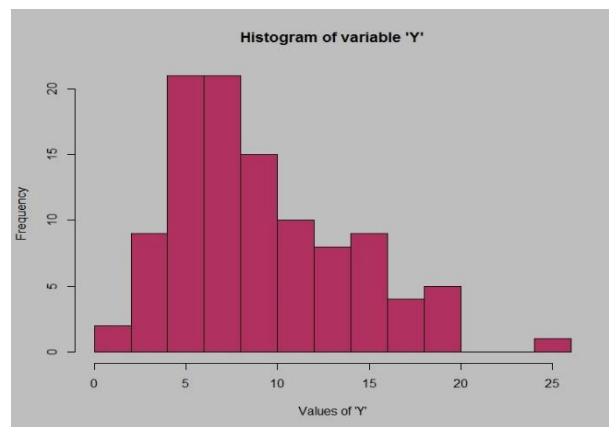
- (i) What is inter-quartile range of this dataset? (Please approximate the numbers) In one line, explain what this value implies.
- (ii) What can we say about the skewness of this dataset?

**(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) From the boxplot given, the Inter Quartile Range maybe estimated to be from 5 to 12 (approximately!). The IQR tells how spread out the "middle" values are; it can also be used to find the outliers.
- (ii) Here the plot is aligned leftward. Which implies this must be a right tailed plot with extreme possible outliers to the right side of the graph. Hence the skewness of the distribution ought to be positive.
- (iii) The case in point pertains only to an extreme point that already had no effect on the boxplot before. With the current development, the point will be relocated to the lower quartile. However, the mean value of the dataset will be susceptible to a considerable change.

**3.**

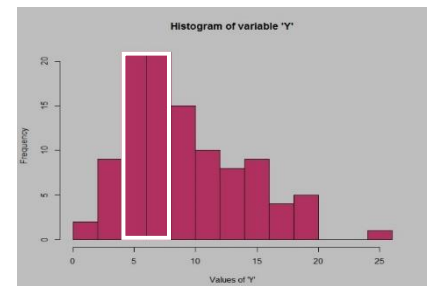


**Answer the following three questions based on the histogram above.**

- (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 of the above dataset is likely to be found where the frequency of occurring is maximum. That means around the highlighted region as marked. Let us say between 4 and 7 (approximately)
- (ii) The given dataset has a positive skewness and is right-tailed. The outliers are more possibly found at the right end of the distribution.



- (iii) Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability. Histograms are preferred to determine the underlying probability

distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.

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.*

Here,

Probability of a call getting misdirected is  $1/200$

So, probability that call does not get misdirected =  $1 - P(E) = 199/200$

Now,

$N = 5$ ,

Probability that at least one among the 5 attempts calls is misdirected,

$P(\text{one in five is misdirected}) = 1 - P(\text{No calls are misdirected in the 5 attempts})$

$$\begin{aligned} P(X) &= 1 - \left(\frac{199}{200}\right)^5 \\ &= 1 - 0.9752487531 \\ &= 0.02475125 \end{aligned}$$

Probability that at least one among the 5 attempts calls is misdirected is 0.02475125

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) In the given data,  $x=2000$  has the highest probability of 0.3. Hence 2000 is the most likely outcome.

(ii)

$$\begin{aligned} P(x>0) &= P(x=1000)+P(x=2000)+P(x=3000) \\ &= 0.2 + 0.3 + 0.1 \\ &= 0.6 \end{aligned}$$

$$\begin{aligned} P(x\leq 0) &= P(x=0)+P(x=-1000)+P(x=-2000) \\ &= 0.2 + 0.1 + 0.1 \\ &= 0.4 \end{aligned}$$

Since  $P(x>0)$  is higher, the venture is likely to succeed.

(iii)

$E(x)$
$-2000 \times 0.1$
$-1000 \times 0.1$
$0 \times 0.2$
$1000 \times 0.2$
$2000 \times 0.3$
$3000 \times 0.1$
$\Sigma = 800$

$E(x) = 800$  is the expected value

(iv) Standard deviation is a great choice.

For the data given,

Standard Deviation,  $\sigma = 1469.6938456699068$