

Topics: Descriptive Statistics and Probability

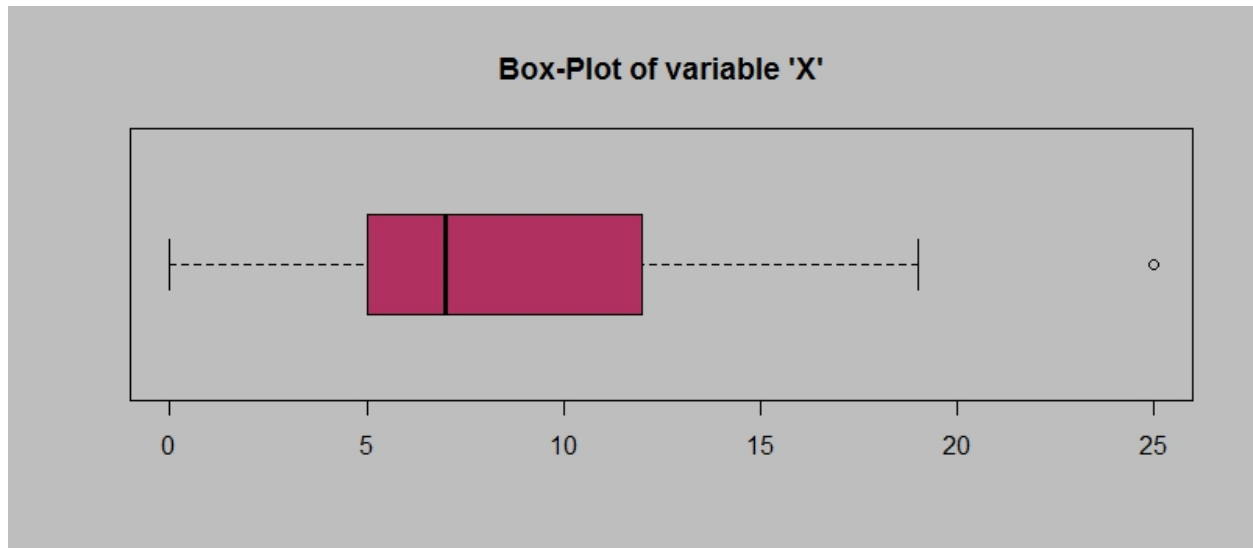
1. Look at the data given below. Plot the data, find the outliers and find out μ, σ, σ^2

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. R CODE:

```
(Copied Data to Excel and saved as Assignment.xls
Data1 = Assignment1 #Saved dataset in excel and imported
boxplot(Data1$`Measure X`) #Plotting boxplot to view outliers
IQR = IQR(Data1$`Measure X`) #Calculating IQR
Q = as.numeric(quantile(Data1$`Measure X`
`)) #For Calculating Q1,and Q3
Q1 = Q[2]
Q3 = Q[4]
l = length(Data1$`Measure X`)
v = Data1$`Measure X`
i=1
while(i<l)
{
if(v[i] > (Q3+1.5*IQR))
print(v[i])
#
0.9136
i=i+1
}
mean = mean(Data1$`Measure X`)
std = sd(Data1$`Measure X`)
var = var(Data1$`Measure X`)
cat("Mean =",mean,"Standard Deviation=",std,"Variance=",var) #Printing Multiple Outputs
# Mean=0.3762933, Standard Deviation=0.2245065, Variance=0.05040318, Outlier=0.9136
```

2.



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

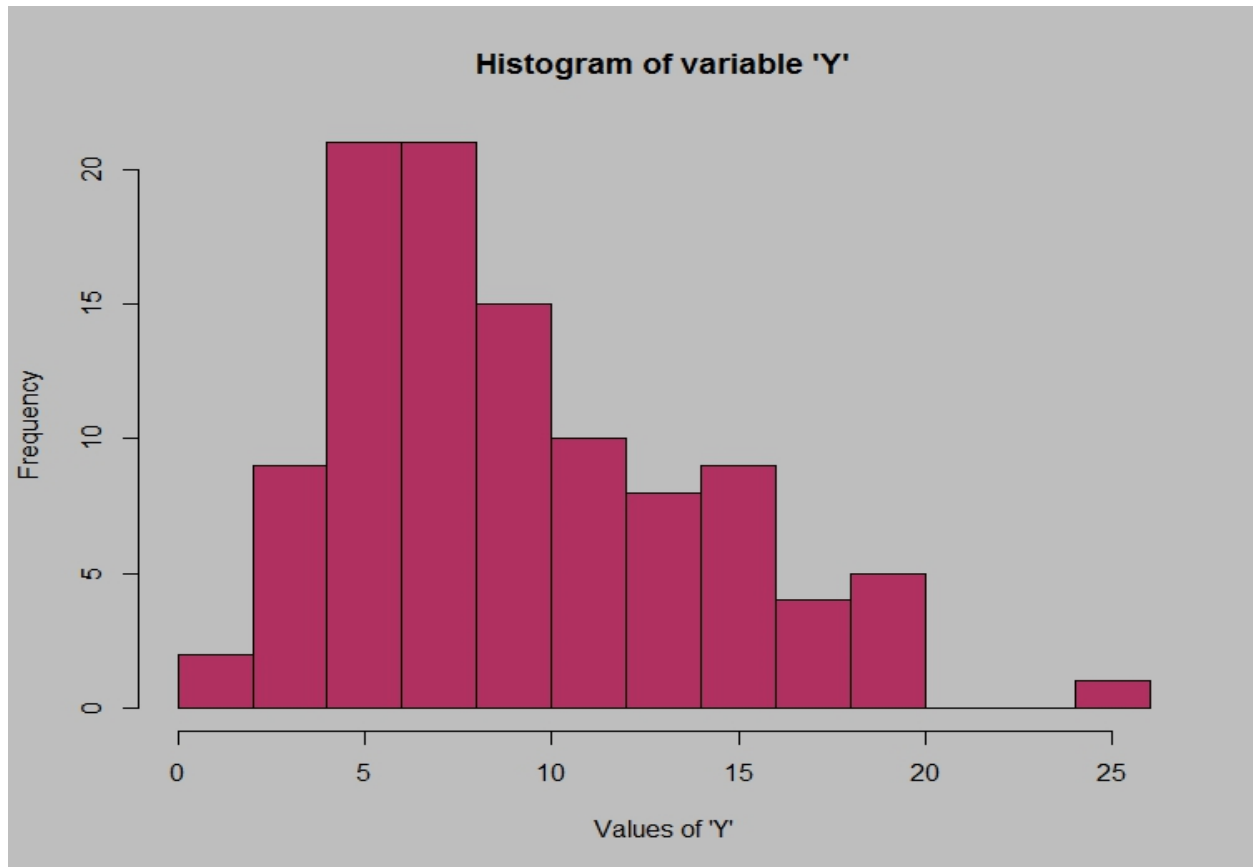
- (i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.
- (ii) What can we say about the skewness of this dataset?
- (iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

Ans. i) $IQR = 12 - 5 = 7$, this represents the range which contains 50% of the data points.

ii) Right Skewed

iii) 2.5 will be not considered an outlier. The boxplot will start from 0 and end at 20 in representation.

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) Mode lies between 4 to 8

ii) Dataset is right skewed.

iii) Median in boxplot and Mode in histogram Histogram provides the frequency distribution so we can see how many times each data point is occurring however boxplot provides the quantile distribution i.e.50% data lies between 5 and 12.Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an 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.)

Ans. $P = 0.028$ (Using MonteCarlo Simulation for 1000 occurrences)

R CODE:

```
d = rep(c("MISS","NO_MISS"),times=c(1,199)) #Creating a sample dataset of 200 outcomes, with only 1 event of Mis
```

```
-
```

```
routing
```

```
#event = sample(d,5,replace=TRUE) #Sample of 5 call scenarios
```

```
event_1000_rep = replicate(1000,sample(d,5,replace=TRUE)) #Repeating experiment 1000 times
```

```
i = 1
```

```
z=0
```

```
while(i<=1000)
```

```
{
```

```
if("MISS" %in% event_1000_rep[,i])
```

```
#Checking if we have even one MISS among the 5 calls
```

```
z=z+1
```

```
i=i+1
```

```
}
```

```
p = z/1000
```

```
cat("Probability
```

```
",p)
```

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) Max. $P = 0.3$ for $P(2000)$ So most likely outcome is 2000.

ii) $P(x > 0) = 0.6$, implies there is a 60% chance that the venture would yield profits or greater than expected returns. $P(\text{Incurring losses})$ is only 0.2. So the venture is likely to be successful.

iii) Weighted average $= x * P(x) = 800$. This means the average expected earnings over a long period of time would be 800. (including all losses and gains over the period of time)