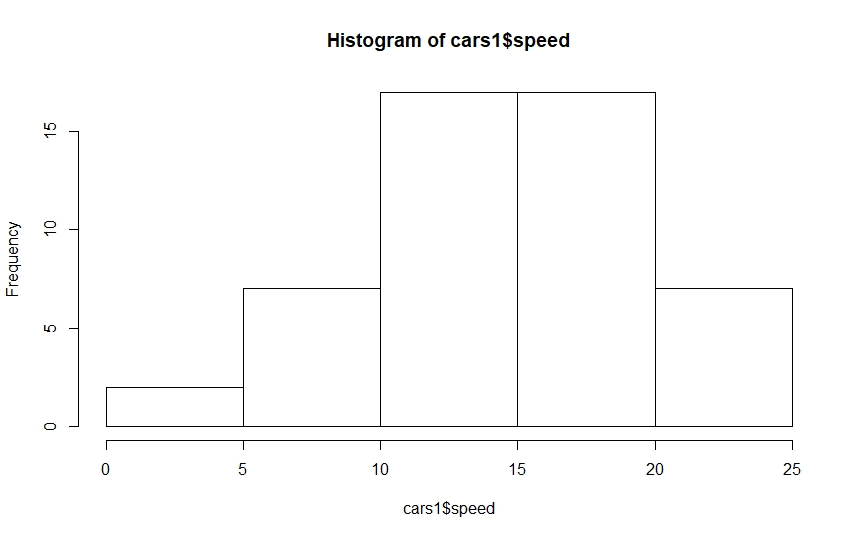
Q1) Calculate Skewness, Kurtosis & draw inferences on the following data

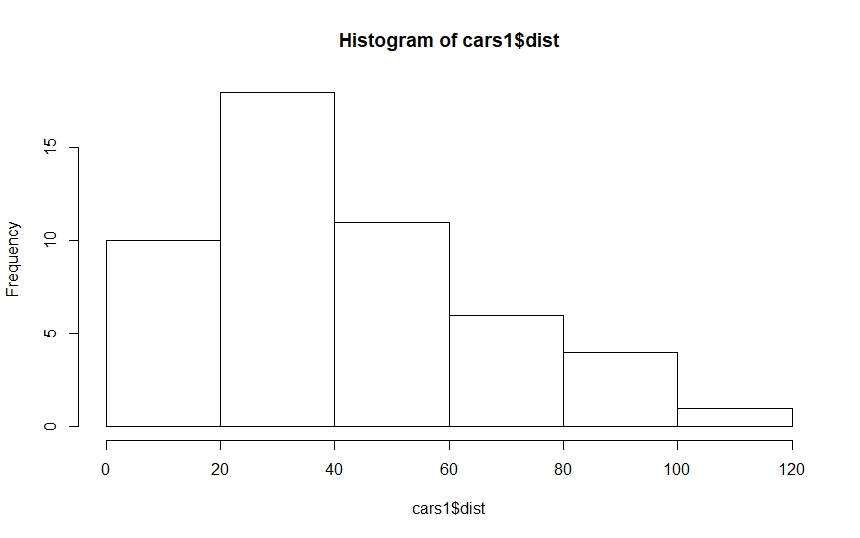
a. Cars speed and distance

****

****

|  |
| --- |
| > skewness(cars1$speed)  [1] -0.1139548  > kurtosis(cars1$speed)  [1] 2.422853 |
|  |
| |  | | --- | |  | |

Inferences: Histogram of Cars data and speed it is Negatively skewed why because it values lies towards Right, and kurtosis of speed peak value is 2, it says data is higher peak at mean from distribution



skewness(cars1$dist)

[1] 0.7824835

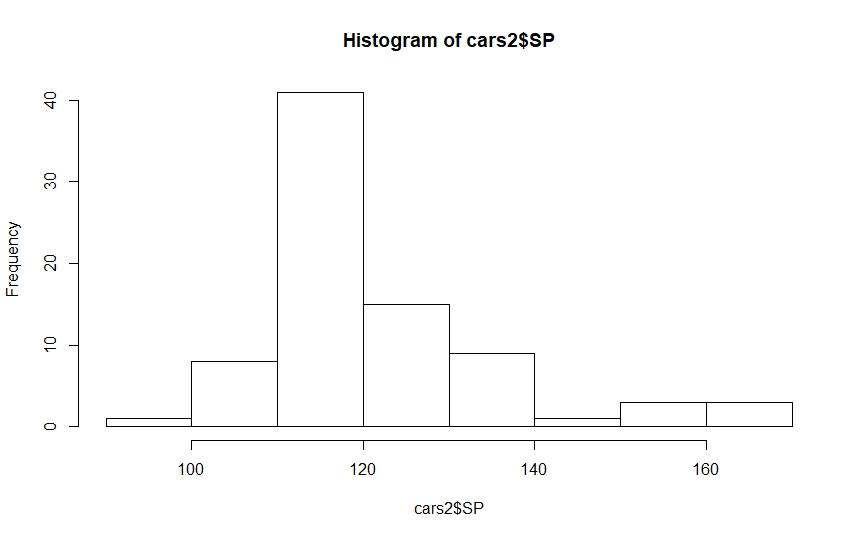
kurtosis(cars$dist)

[1] 3.248019

Inferences: Histogram of Cars distance is Positively skewed. It has data lies in left part and lean towards right, and kurtosis value 3 says it is highly peak from mean

b. Top Speed (SP) and Weight (WT)

****

****

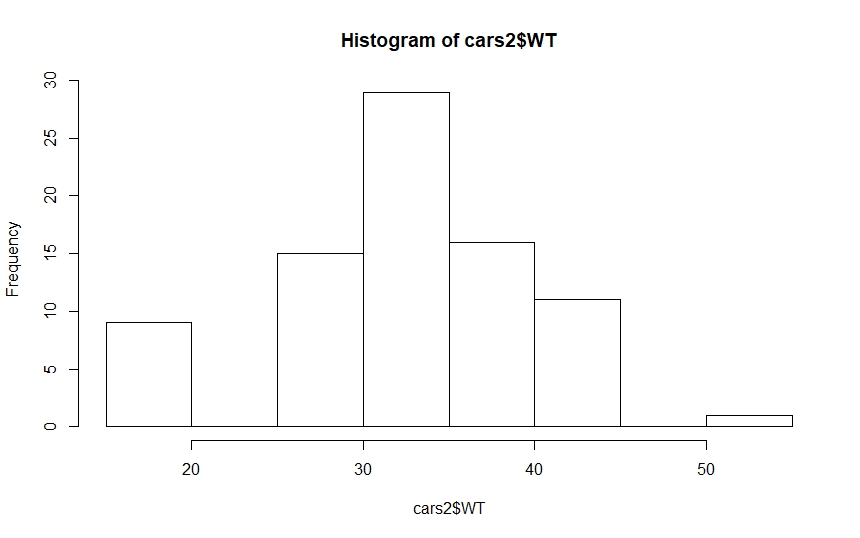
> skewness(cars2$SP)

[1] 1.581454

kurtosis(cars2$SP)

[1] 5.723521

Inference: Cars Data of speed is it is positively skewed towards right. It has higher peak as Kurtosis value is 5 which is higher than mean, it says this distribution



skewness(cars2$WT)

[1] -0.6033099

>

> kurtosis(cars2$WT)

[1] 3.819466

Inference : As Negative value in skewness it shows it is skewed towards left, kurtosis value 3 it implies that it is highly peak from mean.

Q2) Draw inferences about the following boxplot & histogram



Inferences : it says that all the value of this distribution is concentrated at left area and it is positively skewed as it is lies towards right. Mean > median

0

**Inference:** it says that it has lots of outliers concentrated at above and below one outliers which can extremely affect central value. That’s why boxplot above shows the outlier outer part.

**Q3)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Q4)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

mean(x)

[1] 41

> median(x)

[1] 40.5

> var(x)

[1] 25.52941

> sd(x)

[1] 5.052664

1. What can we say about the student marks?

From here students data we can conclude that **mean > median** it means it is **positively skewed.**

Q5) What is the nature of skewness when mean, median of data are equal?

**Symmetric**

Q6) What is the nature of skewness when mean > median?

**Rightly Skewed**

Q7) What is the nature of skewness when median > mean?

**Left Skewed**

Q8) What does positive kurtosis value indicates for a data?

It indicate **Higher peak**

Q9) What does negative kurtosis value indicates for a data?

**Lower peak**

Q10) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

It is negatively skewed towards left

What is nature of skewness of the data?

**LEFT Skewed**

What will be the IQR of the data (approximately)?   
Range is IQR(18-10)

Q11) Comment on the below Boxplot visualizations?



Its central line indicates that it is symmetric distribution. Which means normal Distribution.

Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Q12)



Answer the following three questions based on the boxplot above.

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

**IQR = 13-5= 8**

1. What can we say about the skewness of this dataset?

It is **positively skew.**

1. If it was found that the data point with the value 25 is actually 2.5, how would the new boxplot be affected?

**Value 25 will be Outlier**

Q13)



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Mode Lies in 5 and 10**

1. Comment on the skewness of the dataset.

It is **positively Skewed**

1. Suppose that the above histogram and the boxplot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

Boxplot and Histogram both conveys that Data is **Positively Skewed**