Homework Exercise 1

W203 Statistics for Data Science (Sec 7)

Unit 1 Homework Exercise

Code is submitted in the Gupta_Suhas_HM1.rmd file.

Code run results are submitted in Gupta_Suhas_HW1.html file.

1. Variables in the file *'cars.csv'*:

mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb

2. Summary of mpg variable in '*cars.csv'*:

Summary of mpg variable from the cars data frame	
Mean	19.49
Median	18.70
Minimum	10.40
Maximum	33.90
1st Quartile	15.20
3rd Quartile	21.50

3. Histogram of ‘mpg’ variable is shown in Fig1.

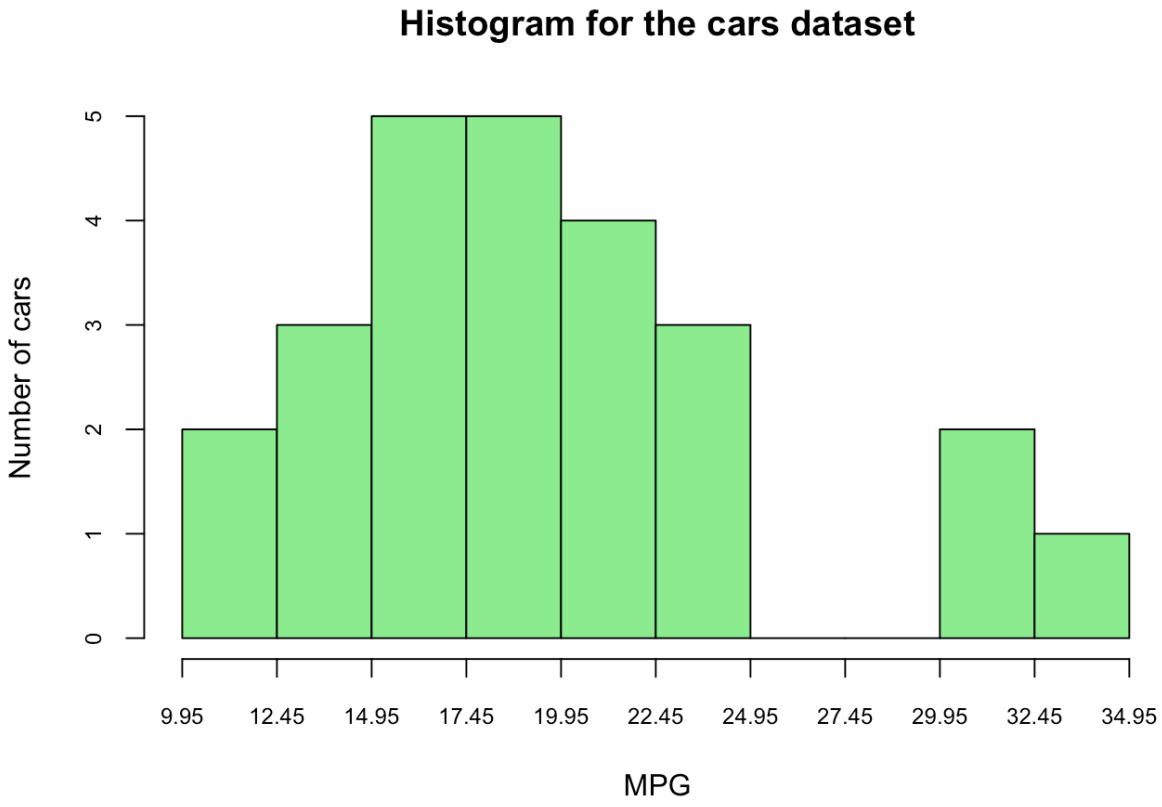

Fig1: Histogram of mpg variable

4. Standard Deviation of MPG = **6.05 mpg**

5. Variance of MPG = **36.572 (mpg)²**

6. The sample variance is denoted by

$$\frac{s^2 = \sum (x_i - x)^2}{(n-1)}$$

The sample standard deviation is the positive square root of the variance

$$s = \sqrt{s^2}$$

The stdDev & Variance of the mpg variable differ because StdDev is defined as the square root of the variance. The standard deviations is defined differently than the variance to provide a metric for easier intuitive understanding of the sample dataset. Taking the square root of the variance lets standard deviation have the same units as the data variable. This helps in rough estimation of the typical size of deviation, of datum points within the sample, from the sample mean.

7. The number of data points in ‘cyl’ variable in cars.csv = **23**
8. Mean of the cal variable = **6.2609**