Test

Huvinesh Rajendran (0340325)

Bachelor of Computer Science

08-10-2021

1 Question 1

Explore and understand the R built-in mtcars data set.

Answer

```
str(mtcars)
```

```
## 'data.frame': 32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

The above code block shows the exploration of the mtcars dataset.This dataset consists of 32 observations with 11 variables which are 'mpg','cyl','disp','hp', 'drat','wt','qsec','vs','am', 'gear', and 'carb'. 'mpg','drat','wt', and 'qsec' are continuous numerical variables because their values can be measured. 'cyl', 'disp', 'hp','vs','am','gear', and 'carb' are discrete numerical variables since they can be counted.

2 Question 2

Describe the mtcars dataset using numerical measures.

Answer

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
var_mpg <- var(mtcars$mpg)
var_cyl <- var(mtcars$cyl)
var_disp <- var(mtcars$disp)
var_hp <- var(mtcars$hp)
var_drat <- var(mtcars$drat)
var_wt <- var(mtcars$wt)
var_qsec <- var(mtcars$qsec)
var_vs <- var(mtcars$vs)
var_am <- var(mtcars$am)
var_gear <- var(mtcars$gear)
var_carb <- var(mtcars$carb)

sd_mpg <- sd(mtcars$mpg)
sd_cyl <- sd(mtcars$cyl)
sd_disp <- sd(mtcars$disp)
sd_hp <- sd(mtcars$hp)
sd_drat <- sd(mtcars$drat)
sd_wt <- sd(mtcars$wt)
sd_qsec <- sd(mtcars$qsec)
```

```r
sd_vs <- sd(mtcars$vs)
sd_am <- sd(mtcars$am)
sd_gear <- sd(mtcars$gear)
sd_carb <- sd(mtcars$carb)
```

Description of mpg
The variance of "mpg" is 36.3241028.
The Standard Deviation of "mpg" is 6.0269481.

Description of cyl
The variance of "cyl" is 3.1895161.
The standard deviation of "cyl" is 3.1895161.

Description of disp
The variance of "disp" is 1.53608×10^4.
The standard deviation of "disp" is 123.9386938.

Description of hp
The variance of "hp" is 4700.8669355.
The standard deviation of "hp" is 68.5628685.

Description of drat
The variance of "drat" is 0.2858814.
The standard deviation of "drat" is 0.5346787.

Description of wt
The variance of "wt" is 0.957379.
The standard deviation of "wt" is 0.9784574.

Description of qsec
The variance of "qsec" is 3.1931661.
The standard deviation of "qsec" is 1.7869432.

Description of vs
The variance of "vs" is 0.2540323.
The standard deviation of "vs" is 0.5040161

Description of am
The variance of "am" is 0.2489919.
The standard deviation of "am" is 0.4989909.

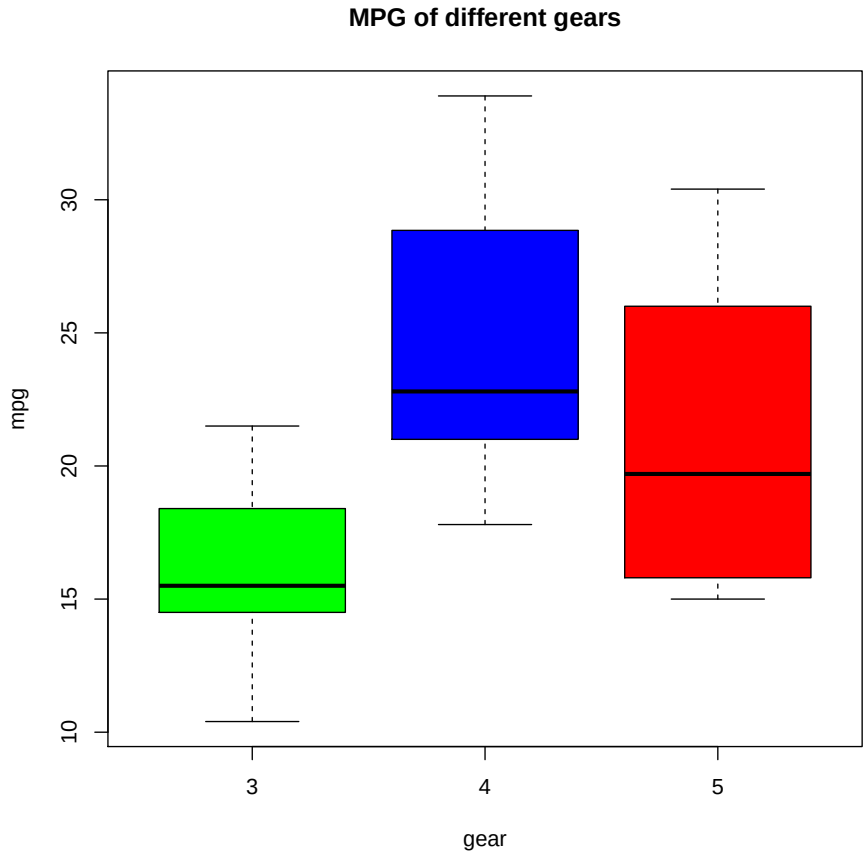
Description of gear
The variance of "gear" is 0.5443548
The standard deviation of "gear" is 0.7378041

Description of carb

The variance of "carb" is 2.608871
The standard deviation of "carb" is 1.6152.

Box plot for the variables "mpg" and "gear":

```
boxplot(mpg ~ gear,data=mtcars,main="MPG of different gears",col= c("green","blue","red"))
```



3 Question 3

Describe the mtcars dataset using tables and graphs. Create a probability distribution table for disp variable.

Answer

```r
disp_max <- max(mtcars$disp)
disp_min <- min(mtcars$disp)
bin <- seq(disp_min,disp_max,by=67)
disp<-cut(mtcars$disp,bin)
freq_table<-transform(table(disp))

transform(freq_table,Relative_Frequency=prop.table(Freq),
          Probability=100*prop.table(Freq))
```

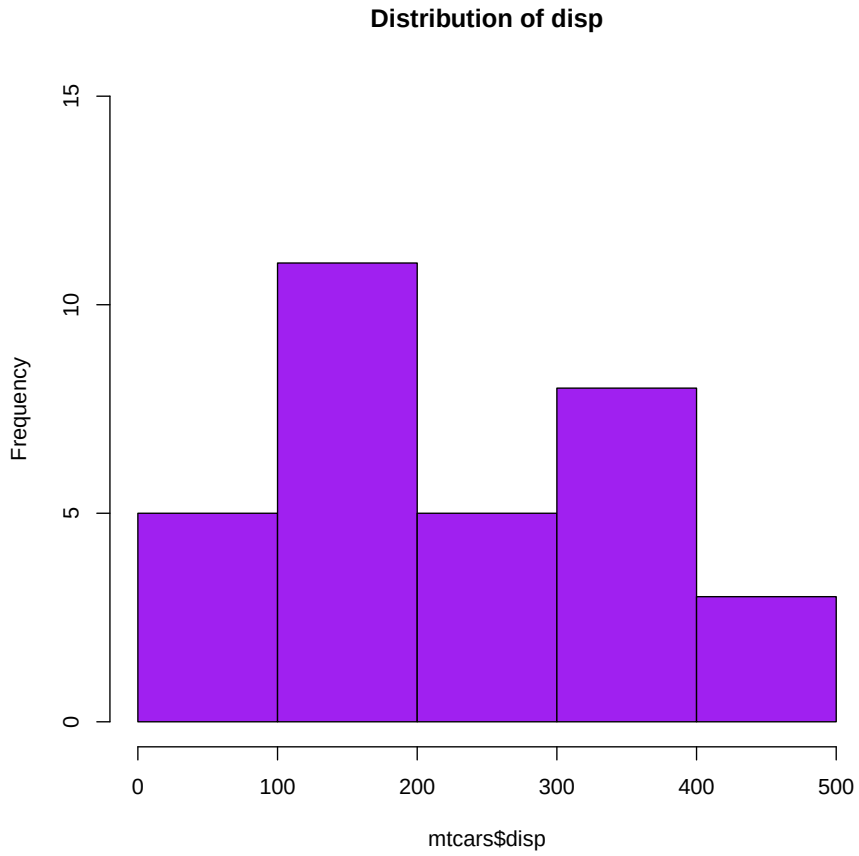
```
##         disp Freq Relative_Frequency Probability
## 1 (71.1,138]    8         0.28571429   28.571429
## 2  (138,205]    7         0.25000000   25.000000
## 3  (205,272]    2         0.07142857    7.142857
## 4  (272,339]    6         0.21428571   21.428571
## 5  (339,406]    5         0.17857143   17.857143
```

4 Question 4

Draw a histogram for the disp variable.

Answer

```r
hist(mtcars$disp,main="Distribution of disp",col="purple",ylim = c(0,15),
     breaks = 4)
```



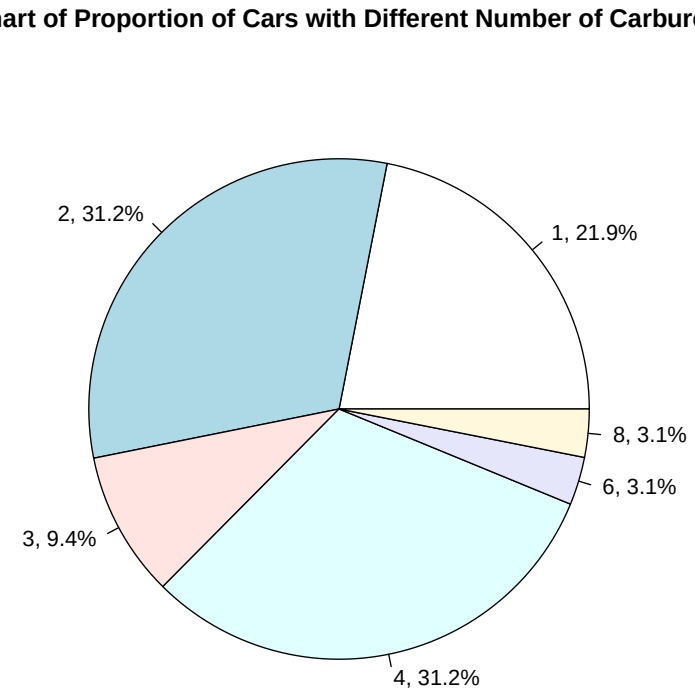
The above histogram shows that the range 100-200 has the highest frequency.

5 Question 5

Draw a piechart for the carb variable.

Answer

```
counts <- table(mtcars$carb)
total <- sum(counts)
proportions <- round((counts*100)/total,digits = 1)
lbls <- paste(names(proportions), ", ", proportions, "%", sep = "")
pie(proportions, labels = lbls, main="Pie Chart of Proportion of Cars with Different Number
```



The above pie chart shows that the caruretors 2 and 4 have the highest distribution with a percentage of 31.2

6 Question 6

Create a contingency table for gear and carb variables.

Answer

The contingeny table for gear and carb variables

```
gear_carb_table <- addmargins(table(mtcars$gear,mtcars$carb))
gear_carb_table
```

```
##
##        1  2  3  4  6  8 Sum
##   3    3  4  3  5  0  0  15
```

```
##   4    4  4  0  4  0  0  12
##   5    0  2  0  1  1  1   5
##   Sum  7 10  3 10  1  1  32
```

This is a contingency table to show the relationship between the variables gear and carb.

7 Question 7

What is the probability when there are 3 gears?

Answer

```
probability_gear_3 <-gear_carb_table[1,7]/32
```

The probability when there are 3 gears is 0.46875. The number of observations with 3 gears is 15. Then it is divided by the total number of observations which is 32.

8 Question 8

What is the probability of 3 gears given that there are 2 carbs.

Answer

```
probability_gear_3_carb_2 <- gear_carb_table[1,2]/32
```

The probability when there are 3 gears and 2 carbs is 0.125. This is because the number of 3 gears and 2 carbs is 4. Then, it is divided by the total number of observations which is 32.

9 Question 9

Proof the Bayesian theorem using the contingency table in Question 6.

Answer

```
probability_carb2 <- gear_carb_table[4,2] / 32

probability_gear3_given_carb2 <-
  probability_gear_3_carb_2/probability_carb2

probability_carb2_given_gear3 <-
  probability_gear_3_carb_2/probability_gear_3
```

The probability of gear 3 given 2 carbs is 0.4.
The probability of 2 carbs given gear 3 is 0.2666667.

```
probability_gear3_given_carb2_2 <-
  (probability_carb2_given_gear3*probability_gear_3)/probability_carb2
```

The answer of the second method of gear 3 given 2 carbs is 0.4.
Hence, the bayesian theorem is proofed.

10 Question 10

Give your solutions in Rnw file, Pdf file and the similarity report.

11 Marking Scheme

Question 1 = 10
Question 2 = 10
Question 3 = 10
Question 4 = 10
Question 5 = 10
Question 6 = 10
Question 7 = 10
Question 8 = 10
Question 9 = 10
Question 10 = 10
Total = 100
Mark= 10