Assignment_4

Zaid

2022-08-14

1. Suppose that the marks secured by 60 students of a class are as follows: 46, 67, 23, 05, 12, 36, 63, 26, 48, 76, 56, 31, 58, 90, 32, 36, 59, 54, 48, 21, 58, 84, 68, 65, 59, 46, 53,64, 57, 65, 53, 38, 58, 26, 43, 45, 66, 74, 16, 86, 43, 36, 66, 46, 58, 36, 64, 58, 45, 76, 74, 48, 64, 58, 50, 58, 95, 56, 66, 44

```
marks=c(46, 67, 23, 05, 12, 36, 63, 26, 48, 76, 56, 31, 58, 90, 32, 36, 59, 54, 48, 21, 58, 84, 68,
65, 59, 46, 53,64, 57, 65, 53, 38, 58, 26, 43, 45, 66, 74, 16, 86, 43, 36, 66, 46, 58, 36,
64, 58, 45, 76, 74, 48, 64, 58, 50, 58, 95, 56, 66, 44)
```

(a) Calculate mean, median, mode, variance and Coefficient of Variation for raw data.

```
marks.mean=mean(marks)
marks.mean
```

```
## [1] 52.53333
```

```
marks.median=median(marks)
marks.median
```

```
## [1] 56
```

```
getmode = function(v) {
   uniqv=unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
marks.mode=getmode(marks)
marks.mode
```

```
## [1] 58
```

```
cv = sd(marks) / mean(marks) * 100
cv
```

```
## [1] 35.70087
```

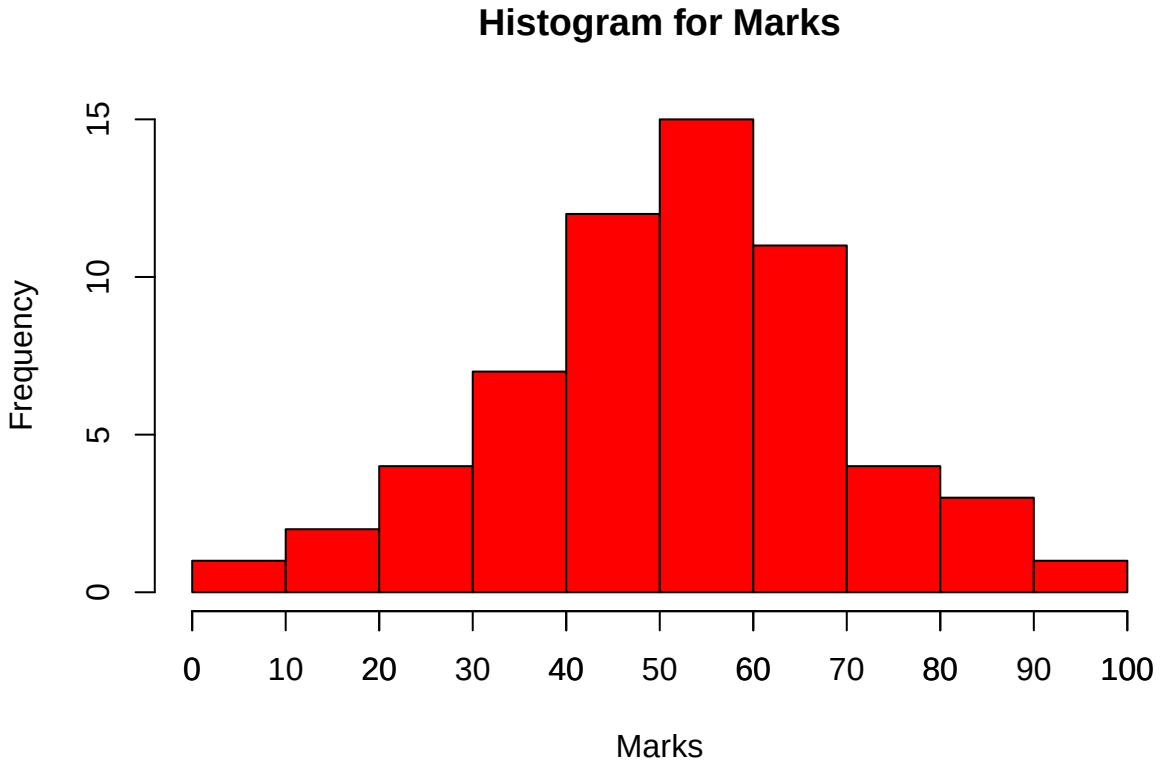
(b) Taking starting point as 0 and common width of the class interval as 10, prepare a frequency table.

```
frequency=table(cut(marks,seq(0,100,10)))
frequency
```

```
##
##   (0,10]  (10,20]  (20,30]  (30,40]  (40,50]  (50,60]  (60,70]  (70,80]
##        1        2        4        7       12       15       11        4
##  (80,90] (90,100]
##        3        1
```

(c) Plot the histogram of the constructed frequency distribution. Give proper x and y axis label and title.

```
hist(marks,main='Histogram for Marks',xlab='Marks',col='red',breaks=10,border='black')
axis(side=1, at=seq(0,100,10))
```

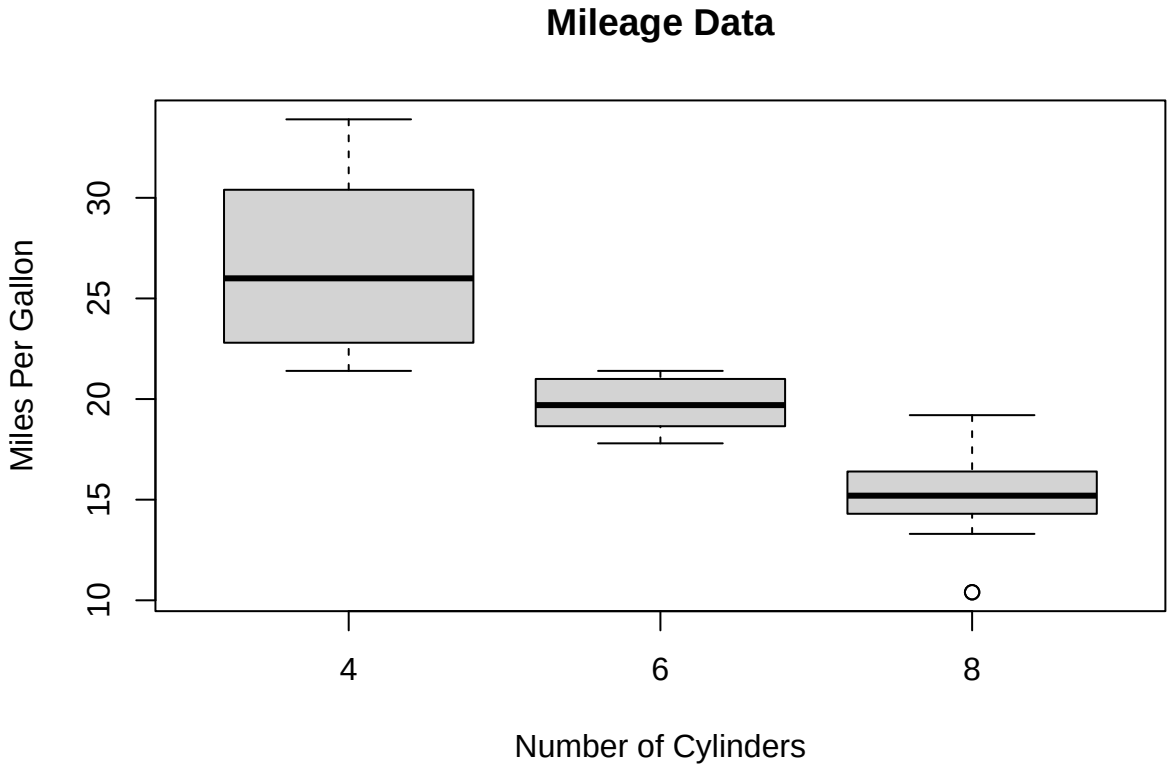


2. Consider any data set and plot the Box and Whisker's Plot and write your interpretation regarding the plot.

```
data= mtcars[,c('mpg','cyl')]
print(head(data))
```

```
##                   mpg cyl
## Mazda RX4         21.0   6
## Mazda RX4 Wag     21.0   6
## Datsun 710        22.8   4
## Hornet 4 Drive    21.4   6
## Hornet Sportabout 18.7   8
## Valiant           18.1   6
```

```
boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders",
   ylab = "Miles Per Gallon", main = "Mileage Data")
```



From the above Box plot, we observe that cars with 4 cylinders give mileage majory in the range of 22 to 30. Similarly, cars with 6 cylinders give mileage majory in the range of 18 to 21 and cars with 8 cylinders give mileage majory in the range of 14 to 16

3. Consider any data set and find the summary statistics using stat.desc function in "pastecs" package, descry function in "summarytools" package and describe function of "psych" package. Interpret the results and write the conclusion.

```
install.packages("pastecs")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library("pastecs")

install.packages("summarytools")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library("summarytools")
```

```
## Warning in fun(libname, pkgname): couldn't connect to display ":0"
```

```
## system might not have X11 capabilities; in case of errors when using dfSummary(), set st_options(use
```

```
install.packages("psych")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library("psych")

data= mtcars[,c('mpg','cyl')]
```

```
stat.desc((data))
```

```
##                      mpg         cyl
## nbr.val       32.0000000  32.0000000
## nbr.null       0.0000000   0.0000000
## nbr.na         0.0000000   0.0000000
## min           10.4000000   4.0000000
## max           33.9000000   8.0000000
## range         23.5000000   4.0000000
## sum          642.9000000 198.0000000
## median        19.2000000   6.0000000
## mean          20.0906250   6.1875000
## SE.mean        1.0654240   0.3157093
## CI.mean.0.95   2.1729465   0.6438934
## var           36.3241028   3.1895161
## std.dev        6.0269481   1.7859216
## coef.var       0.2999881   0.2886338
```

```
descr(data)
```

```
## Descriptive Statistics
## data
## N: 32
##
##                        cyl      mpg
## ----------------- -------- --------
##              Mean     6.19    20.09
##           Std.Dev     1.79     6.03
##               Min     4.00    10.40
##                Q1     4.00    15.35
##            Median     6.00    19.20
##                Q3     8.00    22.80
##               Max     8.00    33.90
##               MAD     2.97     5.41
##               IQR     4.00     7.38
##                CV     0.29     0.30
##          Skewness    -0.17     0.61
##       SE.Skewness     0.41     0.41
##          Kurtosis    -1.76    -0.37
##           N.Valid    32.00    32.00
##         Pct.Valid   100.00   100.00
```

```
describe(data)
```

```
##     vars  n  mean   sd median trimmed  mad  min  max range  skew kurtosis   se
## mpg    1 32 20.09 6.03   19.2   19.70 5.41 10.4 33.9  23.5  0.61    -0.37 1.07
## cyl    2 32  6.19 1.79    6.0    6.23 2.97  4.0  8.0   4.0 -0.17    -1.76 0.32
```