Homework6

Li Zhang

09/27/2021

Exercise 3.16

What is the class of the R built-in data set mtcars? Show the function that you used to check the class.

```r
# the class of R built-in data set mtcars is a data frame.
class(mtcars)
```

```
## [1] "data.frame"
```

Exercise 3.17

For data set mtcars, select a subset which satisfies that the cyl is 6 and mpg greater than 21.2. The subset should not include the variable carb.

```r
y<-mtcars
y.subset <- y[y$cyl == 6 & y$mpg > 21.2, -11]
y.subset
```

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear
## Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3
```

Exercise 3.18

Use aggregate() function to calculate the median of the variables: mpg, disp, hp, and wt across the number of cylinders cyl in the mtcars data set.

```r
# method 1
y<-mtcars
group_median<-aggregate(cbind(y$mpg, y$disp, y$hp, y$wt), list(y$cyl), median)
colnames(group_median)<-c("cyl", "mpg","disp","hp","wt")
group_median
```

```
##   cyl  mpg  disp    hp    wt
## 1   4 26.0 108.0  91.0 2.200
## 2   6 19.7 167.6 110.0 3.215
## 3   8 15.2 350.5 192.5 3.755
```

```r
# method 2
#d<-aggregate(. ~ cyl, mtcars, median)
#d[,c(1,2,3,4,6)]
```

Exercise 3.19

Test if the variables mpg, disp, hp, wt and qsec have significant difference across the three cyl in the R build data set mtcars. Fill the blanks and interpret your conclusion based on the p-values.

```r
aov.fun.car <- function(temx){
m2 <- aov(temx ~ mtcars$cyl)
summary(m2)
}
sub.car <- mtcars[, c(1,3,4,6,7)]
aov.car <- sapply(sub.car, aov.fun.car)
aov.car
```

```
## $mpg
##             Df Sum Sq Mean Sq F value    Pr(>F)
## mtcars$cyl   1 817.71  817.71  79.561 6.113e-10 ***
## Residuals   30 308.33   10.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $disp
##             Df Sum Sq Mean Sq F value    Pr(>F)
## mtcars$cyl   1 387454  387454     131 1.803e-12 ***
## Residuals   30  88731    2958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $hp
##             Df Sum Sq Mean Sq F value    Pr(>F)
## mtcars$cyl   1 100984  100984   67.71 3.478e-09 ***
## Residuals   30  44743    1491
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $wt
##             Df Sum Sq Mean Sq F value    Pr(>F)
## mtcars$cyl   1 18.172 18.1723  47.379 1.218e-07 ***
## Residuals   30 11.507  0.3835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $qsec
##             Df Sum Sq Mean Sq F value    Pr(>F)
## mtcars$cyl   1 34.603  34.603  16.123 0.0003661 ***
## Residuals   30 64.385   2.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# The p-values of the variables mpg, disp, hp, wt and qsec are all close to 0
# Therefore, we reject the null hypothesis.
# the variables mpg, disp, hp, wt and qsec
# have significant difference across the three cyl.
# We can conclude that mpg among three cyls are not the same;
# disp among three cyls are not same;
# hp among three cyls are not same;
# wt among three cyls are not same;
# qsec among three cyls are not same.
```