NAME – ANSH GOEL

REG NO. – 20BCE1798

COURSE NAME – FOUNDATION OF DATA ANALYTICS (FDA)

COURSE CODE: 3505

DATE – 29th JULY, 2022

LAB 8

Use the newsurvey data obtained by cleaning 'na' values in survey data of MASS package to do the following:

1. Find the range of students' age participated in the survey.
2. Break the age range into non-overlapping sub-intervals by defining a sequence of equal distance break points of 10 by rounding the range to nearest integer.
3. Find the distribution of the age range according to the sub-intervals with cut with its right boundary opened. Display it in column form.
4. Which age range of students has mostly participated in the survey.
5. Similarly, find the frequency distribution of Wr.Hnd span and display it in column format.
6. Find the relative frequency of Wr.Hnd and display it by correcting to 3 decimal places.

```
> library(MASS)
> data = survey[complete.cases(survey),]
> data
```

	Sex	Wr.Hnd	NW.Hnd	W.Hnd	Fold	Pulse	Clap	Exer	Smoke	Height	M.I	Age
1	Female	18.5	18.0	Right	R on L	92	Left	Some	Never	173.00	Metric	18.250
2	Male	19.5	20.5	Left	R on L	104	Left	None	Regul	177.80	Imperial	17.583
5	Male	20.0	20.0	Right	Neither	35	Right	Some	Never	165.00	Metric	23.667
6	Female	18.0	17.7	Right	L on R	64	Right	Some	Never	172.72	Imperial	21.000
7	Male	17.7	17.7	Right	L on R	83	Right	Freq	Never	182.88	Imperial	18.833
8	Female	17.0	17.3	Right	R on L	74	Right	Freq	Never	157.00	Metric	35.833
9	Male	20.0	19.5	Right	R on L	72	Right	Some	Never	175.00	Metric	19.000
10	Male	18.5	18.5	Right	R on L	90	Right	Some	Never	167.00	Metric	22.333
11	Female	17.0	17.2	Right	L on R	80	Right	Freq	Never	156.20	Imperial	28.500
14	Female	19.5	20.2	Right	L on R	66	Neither	Some	Never	155.00	Metric	17.500
17	Female	18.0	18.0	Right	L on R	89	Neither	Freq	Never	157.00	Metric	19.333
18	Male	19.4	19.2	Left	R on L	74	Right	Some	Never	182.88	Imperial	18.333
20	Male	21.0	20.9	Right	R on L	78	Right	Freq	Never	177.00	Metric	17.917
21	Male	21.5	22.0	Right	R on L	72	Left	Freq	Never	190.50	Imperial	17.917
22	Male	20.1	20.7	Right	L on R	72	Right	Freq	Never	180.34	Imperial	18.167
23	Male	18.5	18.0	Right	L on R	64	Right	Freq	Never	180.34	Imperial	17.833
24	Male	21.5	21.2	Right	R on L	62	Right	Some	Never	184.00	Metric	18.250
27	Male	21.0	20.7	Right	R on L	90	Right	Some	Never	172.72	Imperial	17.500
28	Male	20.8	21.4	Right	R on L	62	Neither	Freq	Never	175.26	Imperial	18.083
30	Male	19.5	19.5	Right	L on R	79	Right	Some	Never	167.00	Metric	19.250
32	Male	18.8	18.2	Right	L on R	78	Right	Freq	Never	180.00	Metric	17.500
33	Female	17.1	17.5	Right	R on L	72	Right	Freq	Heavy	166.40	Imperial	39.750
34	Male	20.1	20.0	Right	R on L	70	Right	Some	Never	180.00	Metric	17.167
36	Male	22.2	21.0	Right	L on R	66	Right	Freq	Occas	190.00	Metric	18.000
38	Male	19.4	18.5	Right	R on L	72	Neither	Freq	Never	182.50	Metric	17.917
39	Male	22.0	22.0	Right	R on L	80	Right	Some	Never	185.00	Metric	35.500
42	Female	17.8	18.0	Right	R on L	72	Right	Some	Never	154.94	Imperial	17.083
44	Female	20.1	20.2	Right	L on R	80	Right	Some	Never	176.50	Imperial	17.500

1. Find the range of students age participated in the survey

```
>
> ran = range(data$Age)
> ran
[1] 16.917 70.417
>
> |
```

2. Break the age range into non-overlapping sub-intervals by defining a sequence of equal distance break points of 10 by rounding the range to nearest integer.

```
> breaks <-seq(16,76,by=10)
> age_range_brk<-cut(r,breaks,right=TRUE)
> age_range_brk
[1] (16,26] (66,76]
Levels: (16,26] (26,36] (36,46] (46,56] (56,66] (66,76]
```

3. Find the distribution of the age range according to the sub-intervals with cut with its right boundary opened. Display it in column form.

```
> sequence<-seq(16,76,by=10)
> age<-cut(data$Age,sequence,right=FALSE)
> age
  [1] [16,26) [16,26) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
 [16] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [36,46) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26)
 [31] [16,26) [26,36) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
 [46] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
 [61] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
 [76] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
 [91] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26) [16,26)
[106] [66,76) [36,46) [16,26) [36,46) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26)
[121] [16,26) [36,46) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
[136] [26,36) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
[151] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)
[166] [16,26) [16,26) [16,26)
Levels: [16,26) [26,36) [36,46) [46,56) [56,66) [66,76)
> age_freq<-table(age)
> age_freq
age
[16,26) [26,36) [36,46) [46,56) [56,66) [66,76)
    154       9       4       0       0       1
```

```
> age_freq<-table(age)
> age_freq
age
[16,26) [26,36) [36,46) [46,56) [56,66) [66,76)
    154       9       4       0       0       1
> cbind(age_freq)
        age_freq
[16,26)      154
[26,36)        9
[36,46)        4
[46,56)        0
[56,66)        0
[66,76)        1
```

4. Which age range of students has mostly participated in the survey.

```
> print(which.max(age_freq))
[16,26)
     1
```

5. Similarly, find the frequency distribution of Wr.Hnd span and display it in column format.

```
> whs<-data$Wr.Hnd
> range(whs)
[1] 13.0 23.2
> span<-seq(13,24,by=11)
> wr_dist<-table(cut(whs,span,right=TRUE))
> print(cbind(wr_dist))
        wr_dist
(13,24]     167
>
```

6. Find the relative frequency of Wr.Hnd and display it by correcting to 3 decimal places.

```
> options(digits = 3)
> print(wr_dist/nrow(data))

(13,24]
  0.994
>
```

7. Obtain the summary statistics of pulse rate of the students.

```
> summary(data$Pulse)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   35.0    66.8    72.0    74.0    80.0   104.0
>
```

Use the newsurvey data obtained by cleaning 'na' values in survey data of MASS package to do the following:

1. Find the average age of the students participated in the survey.
2. Compute the standard deviation and variance of the height of the students participated in the survey.
3. Compute the quartile of wirte hand span.
4. Find the correlation between write hand span and pulse rate of the students.
5. Find the average age of the students based on how often the student exercises.
6. Find the standard deviation of height of the students under different categories of span of non-writing hand.
7. Obtain the summary statistics of pulse rate of the students.

1. Find the average age of the students participated in the survey.

```
> mean(data$Age)
[1] 20.4
>
```

2. Compute the standard deviation and variance of the height of the students participated in the survey.

```
> var(data$Height)
[1] 99.2
> sd(data$Height)
[1] 9.96
>
```

3. Compute the quartile of wirte hand span.

```
> quantile(data$Wr.Hnd)
  0%  25%  50%  75% 100%
13.0 17.5 18.5 20.0 23.2
>
```

4. Find the correlation between write hand span and pulse rate of the students.

```
> cor(data$Wr.Hnd,data$Pulse)
[1] -0.0138
>
```

5. Find the average age of the students based on how often the student exercises.

```
> tapply(data$Age,data$Exer,mean)
Freq None Some
20.8 21.4 19.8
>
```

6. Find the standard deviation of height of the students under different categories of span of non-writing hand.

```
> tapply(data$Height,data$Nw.Hnd,sd)
  12.5     13   13.5     15   15.4   15.5   15.8     16   16.2   16.4   16.5   16.6   16.7   16.9     17   17.1   17.2   17.3
    NA     NA     NA     NA     NA  5.388  1.414 10.409     NA     NA  6.716     NA  7.905     NA  3.573     NA  8.344  9.115
  17.5   17.6   17.7   17.8   17.9     18   18.1   18.2   18.3   18.4   18.5   18.6   18.8   18.9     19   19.1   19.2   19.5
 5.982  5.096  7.184  7.655     NA 10.662     NA  8.740     NA  9.504  6.542 10.607  4.045     NA  8.085  5.516  8.203  6.769
  19.6   19.7   19.8     20   20.1   20.2   20.3   20.4   20.5   20.7   20.8   20.9     21   21.2   21.4   21.5   21.6     22
 7.718     NA  0.000  8.597     NA 15.638     NA     NA  3.592  6.227     NA     NA  6.850     NA     NA  8.742     NA  4.718
  22.3   22.5   22.6   22.7     23   23.2   23.5
    NA  5.099     NA     NA     NA  0.707     NA
>
```

7. Obtain the summary statistics of pulse rate of the students.

```
> summary(data$Pulse)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   35.0    66.8    72.0    74.0    80.0   104.0
>
```