

**LAB 7 - FOUNDATIONS OF DATA ANALYTICS**  
**MEHER SHRISHTI NIGAM 20BRS1193**

## CODE

```
# Lab 7
# L7+L8
# Meher Shrishti Nigam
# CSE AI + Robotics
# 20BRS1193

# Exp-7 Statistical Analysis of quantitative data
library("MASS")
newsurvey <- survey[rowSums(is.na(survey)) <= 0,]
# newsurvey

#1. Find the range of students' age participated in the survey.
age_range <- array(range(newsurvey$Age))
age_range

#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.
ll <- floor(seq(age_range[1], age_range[2], by=10))
ul <- floor(seq(age_range[1]+10, age_range[2]+10, by=10))
sub_intervals <- data.frame(ll, ul)
sub_intervals
# OR
age_range <- floor(age_range)
age_range
age_range_sub_intervals <- cut(newsurvey$Age, seq(age_range[1], age_range[2]+10, by=10),
right=FALSE)
age_range_sub_intervals

#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.

age_range_sub_intervals_freq <- table(age_range_sub_intervals)
age_range_sub_intervals_freq
freq_age <- cbind(age_range_sub_intervals_freq)
freq_age

#4. Which age range of students has mostly participated in the survey.
freq_age <- cbind(age_range_sub_intervals_freq)
which.max(age_range_sub_intervals_freq)

#5. Similarly, find the frequency distribution of Wr.Hnd span and display it in column
#format.
Wr.Hnd_range <- array(range(newsurvey$Wr.Hnd))
Wr.Hnd_range
Wr.Hnd_range <- floor(Wr.Hnd_range)
```

```
Wr.Hnd_range
```

```
Wr.Hnd_range_sub_intervals <- cut(newsurvey$Wr.Hnd, seq(Wr.Hnd_range[1],  
Wr.Hnd_range[2]+2, by=2), right=FALSE)  
Wr.Hnd_range_sub_intervals
```

```
Wr.Hnd_range_sub_intervals_freq <- table(Wr.Hnd_range_sub_intervals)  
Wr.Hnd_range_sub_intervals_freq
```

```
which.max(Wr.Hnd_range_sub_intervals_freq)
```

```
freq_Wr.Hnd <- cbind(Wr.Hnd_range_sub_intervals_freq)
```

```
freq_Wr.Hnd
```

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

```
Wr.Hnd_range_sub_intervals_rel_freq <-  
Wr.Hnd_range_sub_intervals_freq/(length(Wr.Hnd_range_sub_intervals))  
Wr.Hnd_range_sub_intervals_rel_freq
```

## OUTPUT

**> # Exp-7 Statistical Analysis of quantitative data**

```
> library("MASS")
```

```
> newsurvey <- survey[rowSums(is.na(survey)) <= 0,]
```

**> #1. Find the range of students' age participated in the survey.**

```
> age_range <- array(range(newsurvey$Age))
```

```
> age_range
```

```
[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.**

```
> ll <- floor(seq(age_range[1], age_range[2], by=10))
```

```
> ul <- floor(seq(age_range[1]+10, age_range[2]+10, by=10))
```

```
> sub_intervals <- data.frame(ll, ul)
```

```
> sub_intervals
```

```
ll ul
```

```
1 16 26
```

```
2 26 36
```

```
3 36 46
```

```
4 46 56
```

```
5 56 66
```

```
6 66 76
```

```
> # OR
```

```
> age_range <- floor(age_range)
```

```
> age_range
```

```
[1] 16 70
```

```
> age_range_sub_intervals <- cut(newsurvey$Age, seq(age_range[1], age_range[2]+10, by=10),  
right=FALSE)
```

```
> age_range_sub_intervals
```

```
[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) [16,26) [16,26) [16,26)
```

```
[19] [16,26) [16,26) [16,26) [36,46) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26)  
[16,26) [26,36) [16,26) [16,26) [16,26) [16,26)
```

```
[37] [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) [16,26) [16,26) [16,26)
```

```
[55] [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) [16,26) [16,26) [16,26)
```

```
[73] [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) [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) [66,76) [36,46) [16,26)
```

```
[109] [36,46) [16,26) [16,26) [16,26) [26,36) [16,26) [16,26) [26,36) [16,26) [16,26) [16,26) [16,26)  
[16,26) [36,46) [16,26) [16,26) [16,26) [16,26)
```

[127] [16,26) [16,26) [16,26) [16,26) [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)

[145] [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) [16,26) [16,26) [16,26)

[163] [16,26) [16,26) [16,26) [16,26) [16,26) [16,26)

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.

```
> age_range_sub_intervals_freq <- table(age_range_sub_intervals)
```

```
> age_range_sub_intervals_freq
```

```
age_range_sub_intervals
```

```
[16,26) [26,36) [36,46) [46,56) [56,66) [66,76)
```

```
154     9     4     0     0     1
```

```
> freq_age <- cbind(age_range_sub_intervals_freq)
```

```
> freq_age
```

```
      age_range_sub_intervals_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.

```
> freq_age <- cbind(age_range_sub_intervals_freq)
```

```
> which.max(age_range_sub_intervals_freq)
```

[16,26)

1

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

> **#format.**

> Wr.Hnd\_range <- array(range(newsurvey\$Wr.Hnd))

> Wr.Hnd\_range

[1] 13.0 23.2

> Wr.Hnd\_range <- floor(Wr.Hnd\_range)

> Wr.Hnd\_range

[1] 13 23

> Wr.Hnd\_range\_sub\_intervals <- cut(newsurvey\$Wr.Hnd, seq(Wr.Hnd\_range[1],  
Wr.Hnd\_range[2]+2, by=2), right=FALSE)

> Wr.Hnd\_range\_sub\_intervals

[1] [17,19) [19,21) [19,21) [17,19) [17,19) [17,19) [19,21) [17,19) [17,19) [19,21) [17,19) [19,21)  
[21,23) [21,23) [19,21) [17,19) [21,23) [21,23)

[19] [19,21) [19,21) [17,19) [17,19) [19,21) [21,23) [19,21) [21,23) [17,19) [19,21) [23,25) [21,23)  
[17,19) [17,19) [21,23) [19,21) [17,19) [19,21)

[37] [21,23) [15,17) [19,21) [21,23) [17,19) [19,21) [17,19) [17,19) [17,19) [15,17) [15,17) [17,19)  
[17,19) [17,19) [19,21) [23,25) [17,19) [17,19)

[55] [17,19) [17,19) [19,21) [17,19) [21,23) [19,21) [17,19) [19,21) [17,19) [17,19) [17,19) [19,21)  
[17,19) [19,21) [17,19) [19,21) [17,19) [19,21)

[73] [15,17) [15,17) [17,19) [23,25) [17,19) [21,23) [21,23) [17,19) [19,21) [17,19) [15,17) [17,19)  
[17,19) [15,17) [21,23) [19,21) [15,17) [17,19)

[91] [23,25) [21,23) [19,21) [17,19) [19,21) [21,23) [19,21) [17,19) [21,23) [19,21) [17,19) [17,19)  
[21,23) [13,15) [15,17) [21,23) [17,19) [19,21)

[109] [17,19) [17,19) [17,19) [19,21) [15,17) [17,19) [19,21) [15,17) [19,21) [19,21) [17,19) [17,19)  
[19,21) [19,21) [15,17) [17,19) [19,21) [13,15)

[127] [17,19) [17,19) [17,19) [19,21) [17,19) [17,19) [17,19) [17,19) [19,21) [21,23) [17,19) [17,19)  
[17,19) [15,17) [15,17) [19,21) [17,19) [15,17)

```
[145] [21,23) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19) [17,19)
[23,25) [15,17) [17,19) [17,19) [19,21) [17,19)
```

```
[163] [17,19) [17,19) [17,19) [17,19) [21,23) [17,19)
```

```
Levels: [13,15) [15,17) [17,19) [19,21) [21,23) [23,25)
```

```
> Wr.Hnd_range_sub_intervals_freq <- table(Wr.Hnd_range_sub_intervals)
```

```
> Wr.Hnd_range_sub_intervals_freq
```

```
Wr.Hnd_range_sub_intervals
```

```
[13,15) [15,17) [17,19) [19,21) [21,23) [23,25)
```

```
      2      16      82      41      22      5
```

```
> which.max(Wr.Hnd_range_sub_intervals_freq)
```

```
[17,19)
```

```
      3
```

```
> freq_Wr.Hnd <- cbind(Wr.Hnd_range_sub_intervals_freq)
```

```
> freq_Wr.Hnd
```

```
Wr.Hnd_range_sub_intervals_freq
```

```
[13,15)      2
```

```
[15,17)     16
```

```
[17,19)     82
```

```
[19,21)     41
```

```
[21,23)     22
```

```
[23,25)      5
```

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

```
> #places.
```

```
> Wr.Hnd_range_sub_intervals_rel_freq <-
```

```
Wr.Hnd_range_sub_intervals_freq/(length(Wr.Hnd_range_sub_intervals))
```

```
> Wr.Hnd_range_sub_intervals_rel_freq
```

```
Wr.Hnd_range_sub_intervals
```

```
[13,15) [15,17) [17,19) [19,21) [21,23) [23,25)
```

```
0.01190476 0.09523810 0.48809524 0.24404762 0.13095238 0.02976190
```