

ACADGILD ASSIGNMENT 8.1

1. Use the package -RcmdrPlugin.IPSUR.

`data(RcmdrTestDrive)`

and perform the below operations:

a. Calculate the average salary by gender and smoking status.

b. Which gender has the highest mean salary?

c. Report the highest mean salary.

d. Compare the spreads for the genders by calculating the standard deviation of salary by gender.

```
> summary(RcmdrTestDrive)
      order      race      smoke      gender      salary
Min.   : 1.00 AfAmer: 18 No :134 Female:95 Min.   :11.62
1st Qu.: 42.75 Asian : 8 Yes: 34 Male :73 1st Qu.:15.93
Median : 84.50 Other : 16          Median :17.59
Mean   : 84.50 White :126          Mean   :17.10
3rd Qu.:126.25          3rd Qu.:18.46
Max.   :168.00          Max.   :21.19

      reduction      before      after      parking
Min.   :4.904 Min.   :51.17 Min.   :48.79 Min.   : 1.000
1st Qu.:5.195 1st Qu.:63.36 1st Qu.:62.80 1st Qu.: 1.000
Median :5.501 Median :67.62 Median :66.94 Median : 2.000
Mean   :5.609 Mean   :67.36 Mean   :66.85 Mean   : 2.524
3rd Qu.:5.989 3rd Qu.:71.28 3rd Qu.:70.88 3rd Qu.: 3.000
Max.   :6.830 Max.   :89.96 Max.   :89.89 Max.   :18.000
```

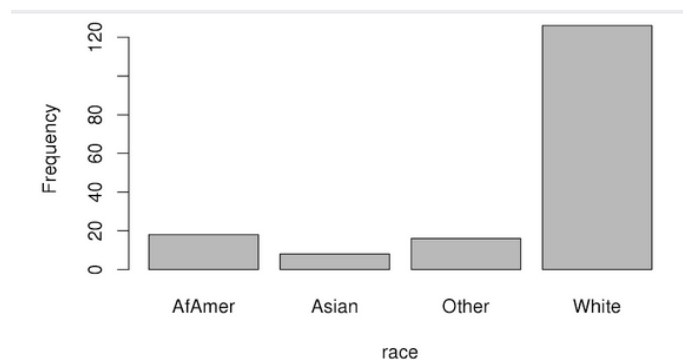
Firstly we need make RACE variable with the table function

```
> table(race)
```

race

AfAmer Asian Other White

18 8 16 126



A) Calculate the average salary by gender and smoking status.

ANSWER:

We can Calculate Average Salary by gender using **tapply** function

```
> x <- tapply(salary, list(gender = gender), mean)
```

```
> x
```

```
gender
```

```
Female    Male
```

```
16.46353 17.93035
```

Using **by** function

```
> by (salary, gender, mean, na.rm = TRUE)
```

```
gender: Female
```

```
[1] 16.46353
```

```
gender: Male
```

```
[1] 17.93035
```

B)Which gender has the highest mean salary?

ANSWER: The Gender with the highest mean salary is **Male**

C)Report the highest mean salary.

ANSWER:

```
Mean( salary [gender = Male])
```

For e.g Highest mean Salary is

```
> x[which(x = max(x))]
```

```
Male
```

```
17.93035
```

D)Compare the spreads for the genders by calculating the standard deviation of salary by gender.

ANSWER:

```
> y <- tapply (salary , list(gender = gender), sd)
```

```
> y
gender
Female    Male
2.122113 1.077183
```

The largest standard deviation is approximately 2.12 which was attained by Female gender.