



## SESSION 8: Exploratory Data Analytics

### Assignment 2

```
1. library(RcmdrPlugin.IPSUR)
```

```
data(RcmdrTestDrive)
```

```
install.packages("RcmdrPlugin.IPSUR")
```

```
install.packages("rlang")
```

```
install.packages("car")
```

```
library(rlang)
```

```
library(Rcmdr)
```

```
library(RcmdrMisc)
```

```
library(RcmdrPlugin.IPSUR)
```

```
library(sandwich)
```

```
library(effects)

library(car)

data("RcmdrTestDrive")

data(BloodPressure)

View(RcmdrTestDrive)

View(BloodPressure)
```

Perform the below operations: -

a. Compute the measures of central tendency for salary and reduction which variable has highest center?

```
> #Find the measures of central tendency for salary and reduction for Salary
> library(RcmdrPlugin.IPSUR)
> x<- c(mean(RcmdrTestDrive$salary), median(RcmdrTestDrive$salary))
> x
[1] 724.5164 710.1500

> #for reduction
> y<- c(median(RcmdrTestDrive$reduction), mean(RcmdrTestDrive$reduction))
> y
[1] 139.500 223.631

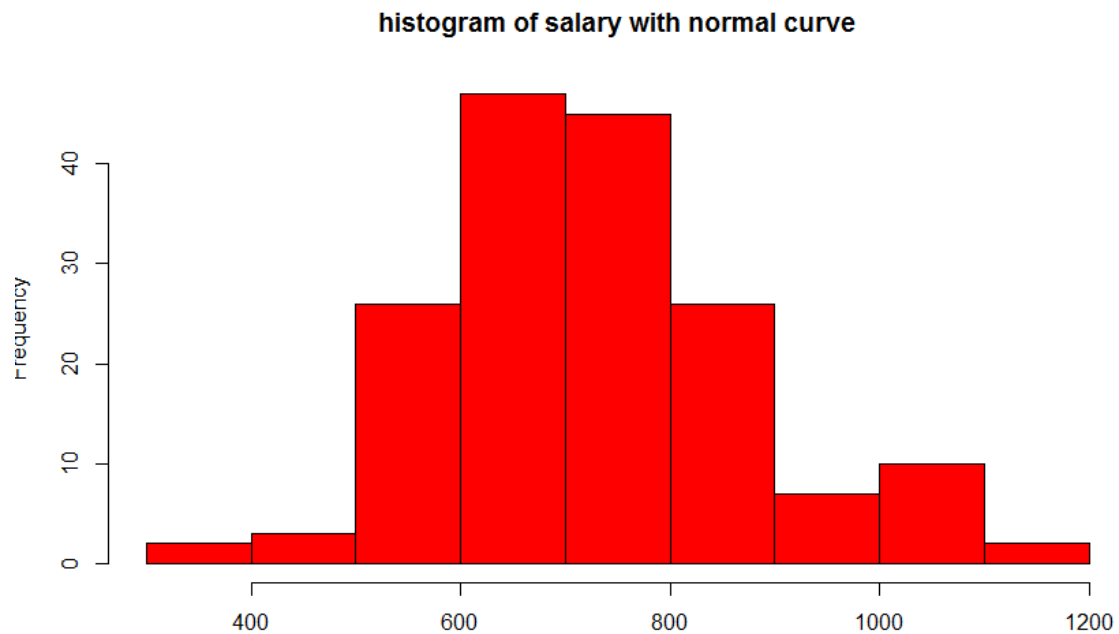
> #now since we are looking for variable which has highest center
> #we can check for this by plotting histogram or
> #by checking kurtosis which describes the amount of peakedness of a distribution.
> #install packages("psych")
> library(psych)

> kurtosi(RcmdrTestDrive$salary)
[1] 0.2006576

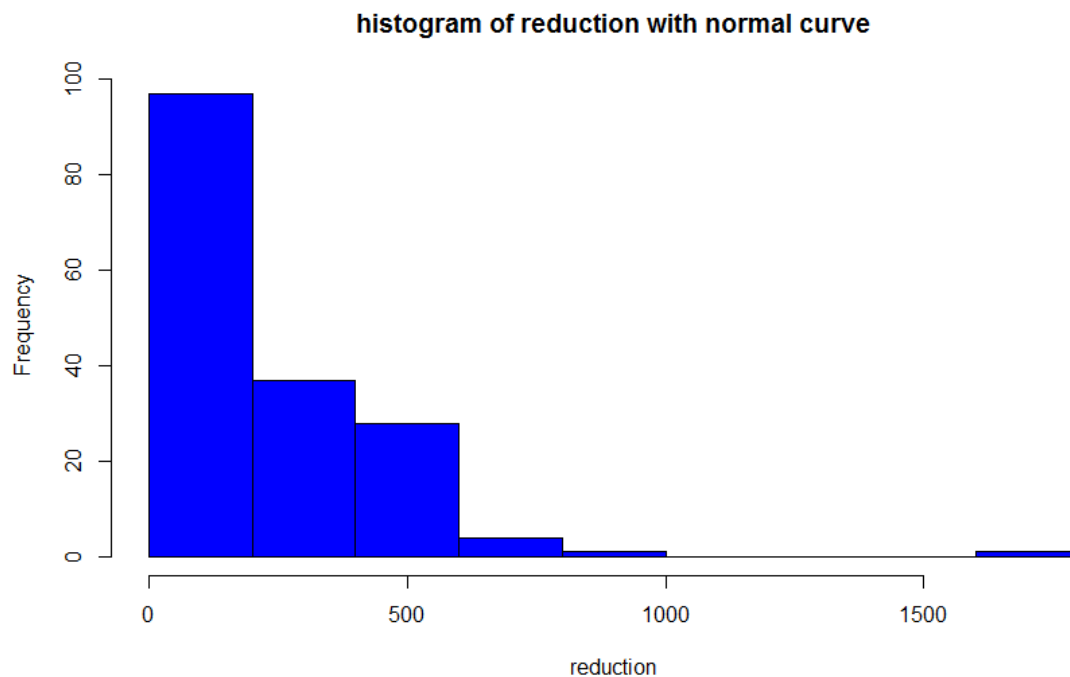
> kurtosi(RcmdrTestDrive$reduction)
[1] 10.01655

> #thus we can see variable reduction has more kurtosis thus more peaked hence more highest center
> #or by plotting histogram we can also check that
>
```

```
> x<-RcmdrTestDrive$salary  
> h<- hist(x,breaks = 10,col = "red",xlab = "salary",main= "hi stogram of s  
alary wi th normal curve")
```



```
> y<-RcmdrTestDrive$reduction  
> h<- hist(y,breaks = 10,col = "blue",xlab = "reduction",main= "hi stogram  
of reduction wi th normal curve")
```



```

> #however as reduction is not purely continuous hence for center we cant see peak of this in from center
> #in that manner salary is more peaked from center as it is purely continuous
> #however variable reduction is more peaked if we talk about the peakedness from whole data
> #by seeing histogram overall as compare to salary variable

```

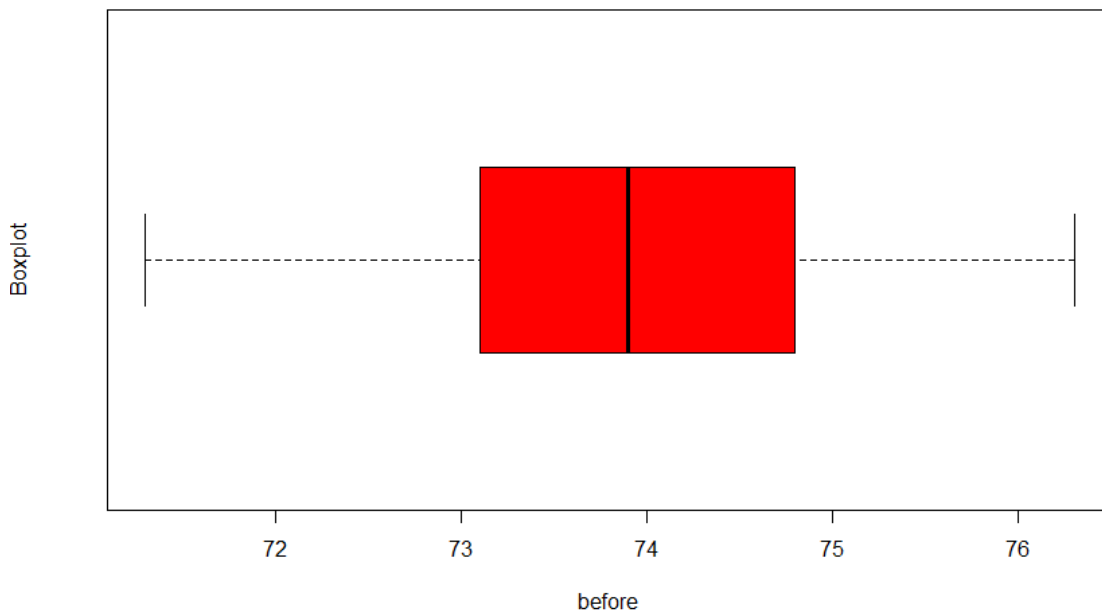
## b. Which measure of center is more appropriate for before and after?

```

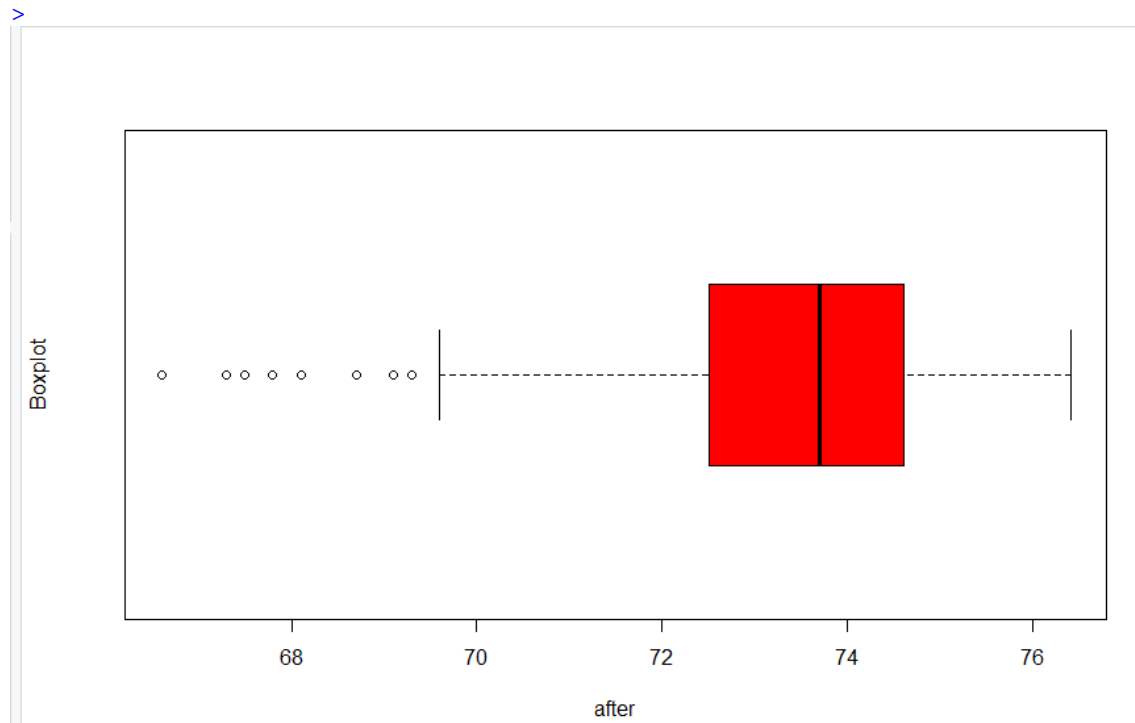
> #If the distribution is fairly symmetric then the mean and median
> #should be approximately the same
>
>
> #by boxplot we can check for median where it lies
> boxplot(RcmdrTestDrive$before, horizontal = T, col = "red", xlab="before", ylab="Boxplot")
> #normal distributed
>

```

---



```
> boxplot(RcmdrTestDrive$after, horizontal = T, col = "red", xlab="after", ylab="Boxplot")
> #left skewed as the data is assymetrical distributed
```



```
> #if we check the skewness of variables
> skew (RcmdrTestDrive$before)
[1] -0.03510369

> skew (RcmdrTestDrive$after)
[1] -1.164056

> #after more negative so data more on right side as compare to before variable
>
> #thus, the median would likely be a good choice and it is more appropriate
> #ps: dots in plots are outliers
```