Missing Value Handling

Example: Suppose a telecom company wants to analyze the performance of its circles based on the following parameters
1. Current Month's Usage
2. Last 3 Month's Usage
3. Average Recharge
4. Projected Growth

The data set is given Missing_Value_Telecom.csv

Example: Read data and variables to R
> mydata = Missing_Values_Telecom
>cmusage = mydata[,2]
> l3musage = mydata[,3]
> avrecharge = mydata[,4]

Option 1: Discard all records with missing values
>newdata = na.omit(mydata)
>write.csv(newdata," ----------")

Option 2: Replace the missing values with variable mean, median, etc
Replacing the missing values with mean
Compute the means excluding the missing values
>cmusage_mean = mean(cmusage, na.rm = TRUE)
>l3musage_mean = mean(l3musage_mean, na.rm = TRUE)
> l3musage_mean = mean(l3musage, na.rm = TRUE)
> avrecharge_mean = mean(avrecharge, na.rm = TRUE)

Replace the missing values with mean
> cmusage[is.na(cmusage)]=cmusage_mean
> l3musage[is.na(l3musage)]= l3musage_mean
>avrecharge[is.na(avrecharge)]=avrecharge_mean

Making the new file
> mynewdata = cbind(cmusage, l3musage, avrecharge, mydata[,5],mydata[,6])
> write.csv(mynewdata, "----------")

DATA MERGING

Exercise: The data of 30 customers on credit card usage in INR1000 is given in CC_Usage.txt. Similarly the user profile namely gender (1: male, 2: female) and whether they have done shopping or banking (1: yes , 2: no) with credit card are given in cc_Profile.csv. Can you merge the two files into a single data set?

Read the files

```
>myprofile = CC_Profile
> myusage = CC_Usage Merge the files by “ID” field
>mydata = merge(myprofile, myusage, by = “ID”)
```

DATA APPEND

Exercise: The data on user profile of customers whom are included in the previous mailing campaign is compiled into two files namely classification1.csv and classification2.txt. Can you append the second data set with the first one and store the new data set in a new file?

Read the files

```
>class1 = Classification1
> class2 = Classification2
```

Append class1 with class2

```
>mydata = rbind(class1, class2)
```

RANDOM SAMPLING

Example: Take a sample of size 60 (10%) randomly from the data given in the file bank-data.csv and save it as a new csv file?

Read the files

```
>mydata = bank-data
> mysample = mydata[sample(1:nrow(mydata), 60, replace = FALSE),]
>write.csv(mysample,"------ ")
```

RANDOM SAMPLING

Example: Split randomly the data given in the file bank-data.csv into sets namely training (75%) and test (25%) ?

Read the files

```
>mydata = bank-data
>sample = sample(2, nrow(mydata), replace = TRUE, prob = c(0.75, 0.25))
> sample1 = mydata[sample ==1, ]
> sample2 = mydata[sample ==2,]
```