**Practice for Learning**

**Data analysis using R**

1. Assign 10 values to a variable. Find its length, sum and mean.
2. Using sequence function to generate the numbers, increment by 10.
3. Install a package ***insurance Data***
4. Load this package
5. From this package, load the data ***SingaporeAuto***
   1. Hint: You may have to use a function ***data( ) or get(data( ))***
6. Find out the nature of the variables, check for NA’s, size and shape
7. Extract first 10 and 40 to 70 cases with first 5 and 11, 13, 15th variables
8. Create new variable based on the
   1. variable Exp\_weights
      1. If Exp\_weights is negative, then newvariable is Low else High
   2. variable NCD

NCD New Variable

0 NIL

10 and 20 Level1

30 Level2

Else Level3

1. Create a subset that has only Clm\_Count = 0 and has only numeric variables
2. Find the counts of
   1. SexInsured
   2. VehicleType
   3. SexInsured and VehicleType
   4. VehicleType and SexInsured
   5. Exp\_weights < 0.75 and VehicleType
3. Draw a scatter plot – EXP\_WEIGHTS

library(dplyr)

library(ggplot2)

**## q.no.1**

a=c(10,20,30,40,24,35,36,67,90,99)

length(a)

sum(a)

mean(a)

**## q.no.2**

seq(from=100,to=2000,by=10)

seq(100,2000,100)

seq.int(from=100,to=2000,length.out=100)

**## q.no.3,4,5**

library(insuranceData)

data("SingaporeAuto")

**## q.no.6**

data=SingaporeAuto

str(data)

is.na(data)

dim(data)

nrow(data)

ncol(data)

## q.no.7

data[c(1:10,40:70),c(1:5,11,13,15)]

**## q.no.8a**

data=data %>% mutate(EXP=case\_when(Exp\_weights<0~"LOW",

TRUE~"HIGH"))

data %>% group\_by(data$EXP) %>% summarise(exp=n())

**## q.no.8b**

data=data %>% mutate(N1=case\_when(NCD==0~"NIL",

NCD==10&NCD==20~"LEVEL1",

NCD==30~"LEVEL2",

TRUE~"LEVEL3"))

data %>% group\_by(data$N1) %>% summarise(N1=n())

**## q.no.9**

data1=data %>% filter(Clm\_Count==0) %>% select\_if(is.numeric)

**## q.no.10**

data %>% group\_by(data$SexInsured) %>% summarise(SI=n())

data %>% group\_by(data$VehicleType) %>% summarise(VT=n())

data %>% group\_by(SexInsured,VehicleType)%>% summarise(SVT=n())

data %>% group\_by(VehicleType,SexInsured)%>% summarise(SVT=n())

data %>% group\_by(data$Exp\_weights<0.75,VehicleType)%>% summarise(EXPVT=n())

**## Q.NO.11**

ggplot(data,aes(x=1:nrow(data),y=Exp\_weights))+geom\_point()

ggplot(data,aes(x=VehicleType,fill=SexInsured))+geom\_bar()+

scale\_fill\_manual(values=c("orange","green","red"))+

geom\_text(aes(label=..count..),stat="count",

position=position\_stack(vjust=0.5),

size=4,color='darkred',fontface="bold")