# RMSC4004 Q18

Chan Sze Yuen Syngiene 1155127616

02/04/2021

**18**

We use the following R code for ploting the histograms:

```r
data=read.csv("D:/CUHKZOOMNOTESANDSOURCE/RMSC4004/AUTOBIsim.csv") #load the data
head(data) #check whether the data is loaded sucessfully
```

```
##   ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE LOSS
## 1  1        1      2       2        2        1    1.8 6.38
## 2  2        2      2       1        2        1    5.6 0.95
## 3  3        1      1       2        2        1    1.4 1.88
## 4  4        2      1       1        2        1    5.4 0.11
## 5  5        1      2       2        2        1    0.9 6.95
## 6  6        2      2       2        2        1    2.3 0.15
```

```r
dataCSEX1=data[data[,"CLMSEX"]==1,] #extract data with "CLMSEX"=1
head(dataCSEX1) #checking
```

```
##    ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE LOSS
## 3   3        1      1       2        2        1    1.4 1.88
## 4   4        2      1       1        2        1    5.4 0.11
## 9   9        2      1       2        2        1    8.2 0.10
## 11 11        1      1       1        2        1    0.6 1.22
## 13 13        1      1       1        2        1    0.5 6.31
## 14 14        1      1       2        2        1    3.8 0.57
```
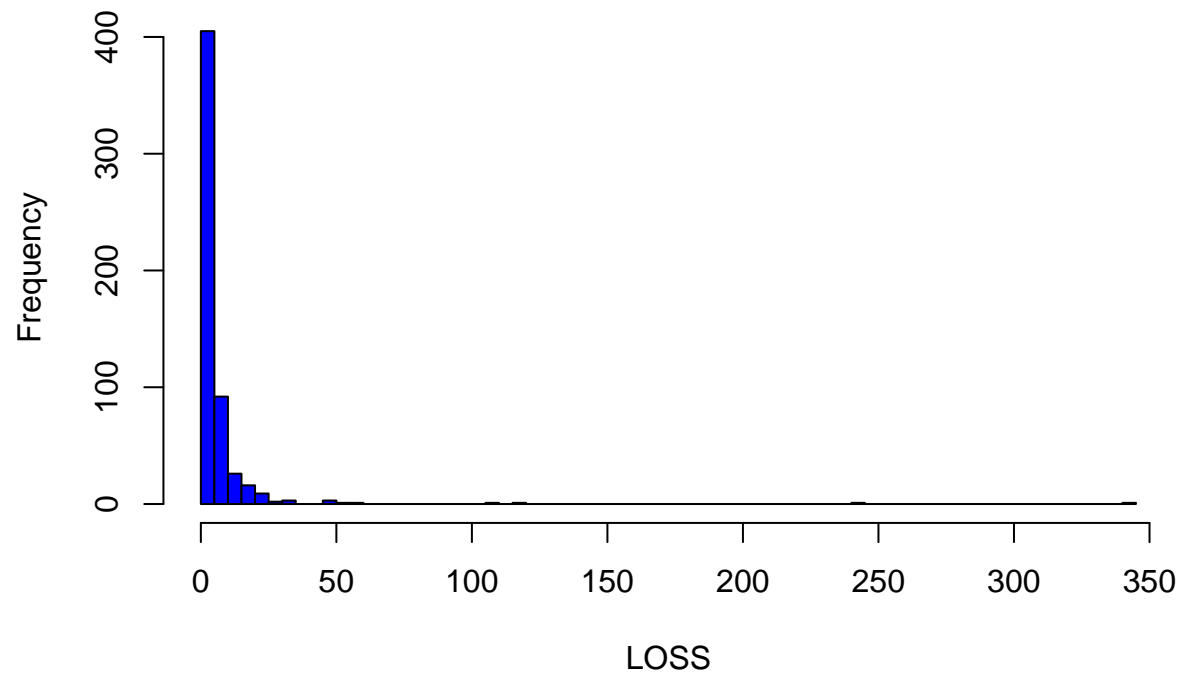
```r
dataCSEX2=data[data[,"CLMSEX"]==2,] #extract data with "CLMSEX"=2
head(dataCSEX2) #checking
```

```
##   ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE  LOSS
## 1  1        1      2       2        2        1    1.8  6.38
## 2  2        2      2       1        2        1    5.6  0.95
## 5  5        1      2       2        2        1    0.9  6.95
## 6  6        2      2       2        2        1    2.3  0.15
## 7  7        1      2       1        2        1    0.6  6.84
## 8  8        1      2       2        2        2    9.3 15.24
```
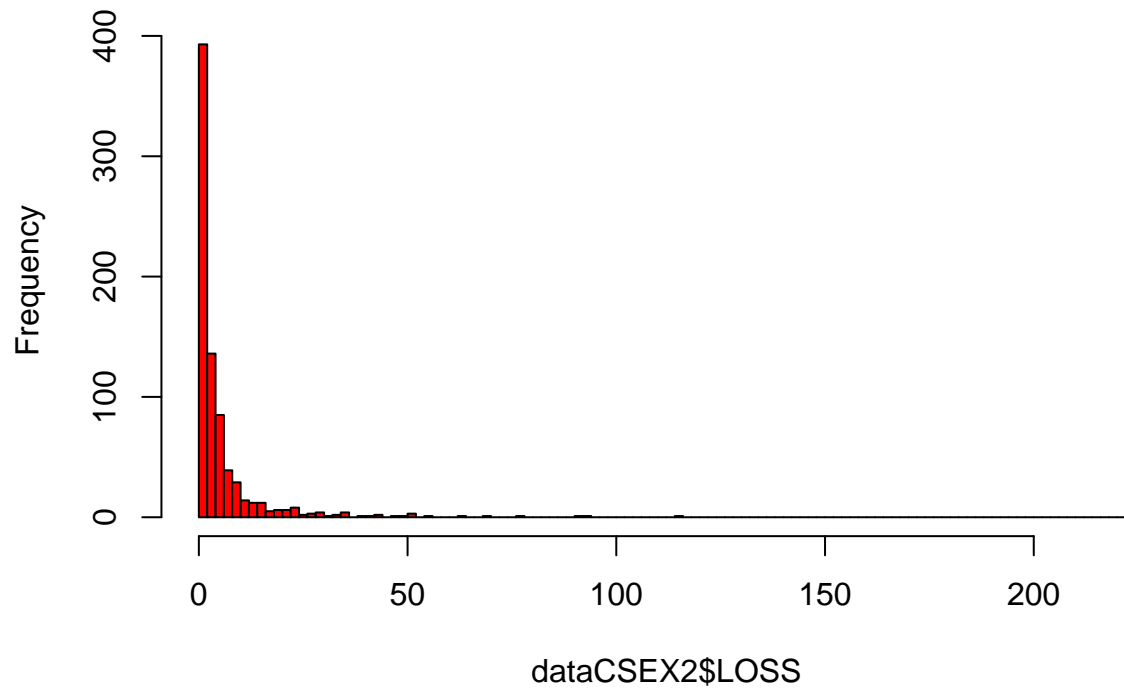
```
p1=hist(dataCSEX1$LOSS,breaks=100,col="blue",xlab="LOSS")
```
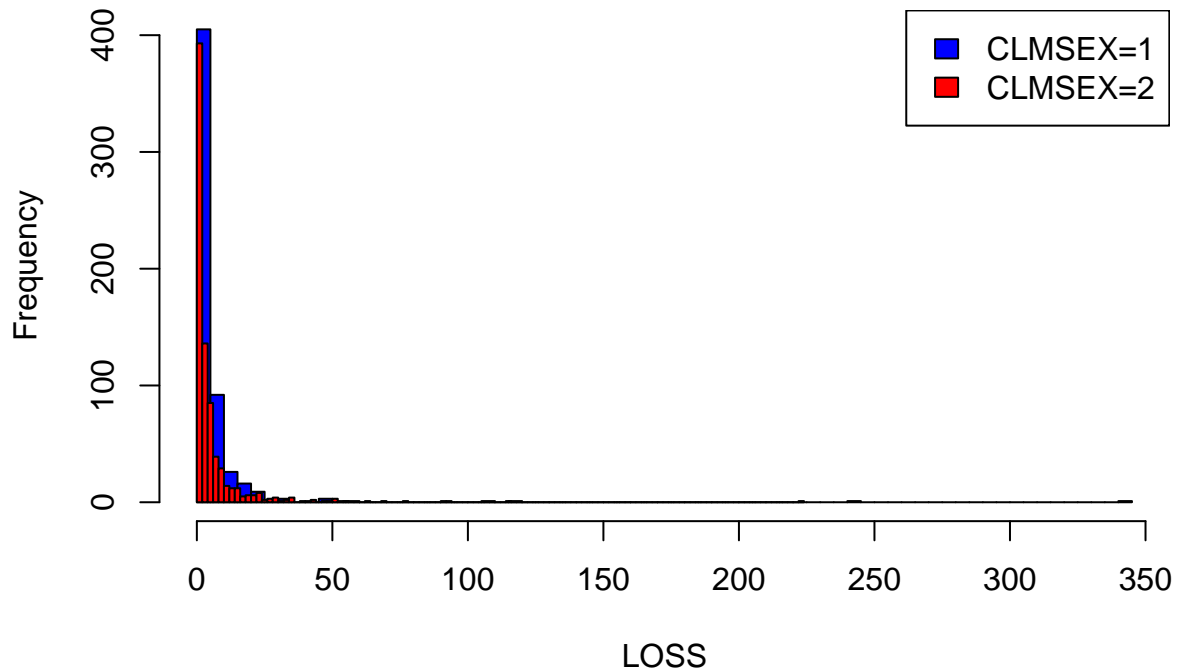
## Histogram of dataCSEX1$LOSS



```
p2=hist(dataCSEX2$LOSS,breaks=100,col="red")
```

## Histogram of dataCSEX2$LOSS



```
plot(p1,col="blue",xlab="LOSS",main="Histogram of LOSS with different CLMSEX")
plot(p2,col="red",add=T)
legend("topright", c("CLMSEX=1", "CLMSEX=2"), fill=c("blue", "red"))
```

## Histogram of LOSS with different CLMSEX



From the plot we can see these two loss distribution have the same distribution with very likely to be exponential distribution, however the loss with CLMSEX=1 one seems having larger tail than the loss with CLMSEX=2 one since most of the extreme value of loss belongs to CLMSEX=1 one. So the answer is they are different (should use two different exponential to model).

For the Histograms of LOSS and different SEATBELT:

```
data=read.csv("D:/CUHKZOOMNOTESANDSOURCE/RMSC4004/AUTOBIsim.csv") #load the data
head(data) #check whether the data is loaded sucessfully
```

```
##   ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE LOSS
## 1  1        1      2       2        2        1    1.8 6.38
## 2  2        2      2       1        2        1    5.6 0.95
## 3  3        1      1       2        2        1    1.4 1.88
## 4  4        2      1       1        2        1    5.4 0.11
## 5  5        1      2       2        2        1    0.9 6.95
## 6  6        2      2       2        2        1    2.3 0.15
```

```
dataSEATBELT1=data[data[,"SEATBELT"]==1,] #extract data with "SEATBELT"=1
head(dataSEATBELT1) #checking
```

```
##   ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE LOSS
## 1  1        1      2       2        2        1    1.8 6.38
## 2  2        2      2       1        2        1    5.6 0.95
## 3  3        1      1       2        2        1    1.4 1.88
## 4  4        2      1       1        2        1    5.4 0.11
```

4

```
## 5   5          1        2          2          2           1      0.9 6.95
## 6   6          2        2          2          2           1      2.3 0.15
```

```r
dataSEATBELT2=data[data[,"SEATBELT"]==2,] #extract data with "SEATBELT"=2
head(dataSEATBELT2) #checking
```

```
##          ID ATTORNEY CLMSEX MARITAL CLMINSUR SEATBELT CLMAGE  LOSS
## 8        8        1      2       2        2        2    9.3 15.24
## 42      42        1      2       2        2        2    9.7  3.48
## 79      79        1      1       1        2        2    1.3 15.80
## 82      82        1      2       2        2        2    5.3 91.18
## 116    116        2      1       1        2        2    1.3  4.63
## 224    224        2      1       1        2        2    0.6  9.69
```
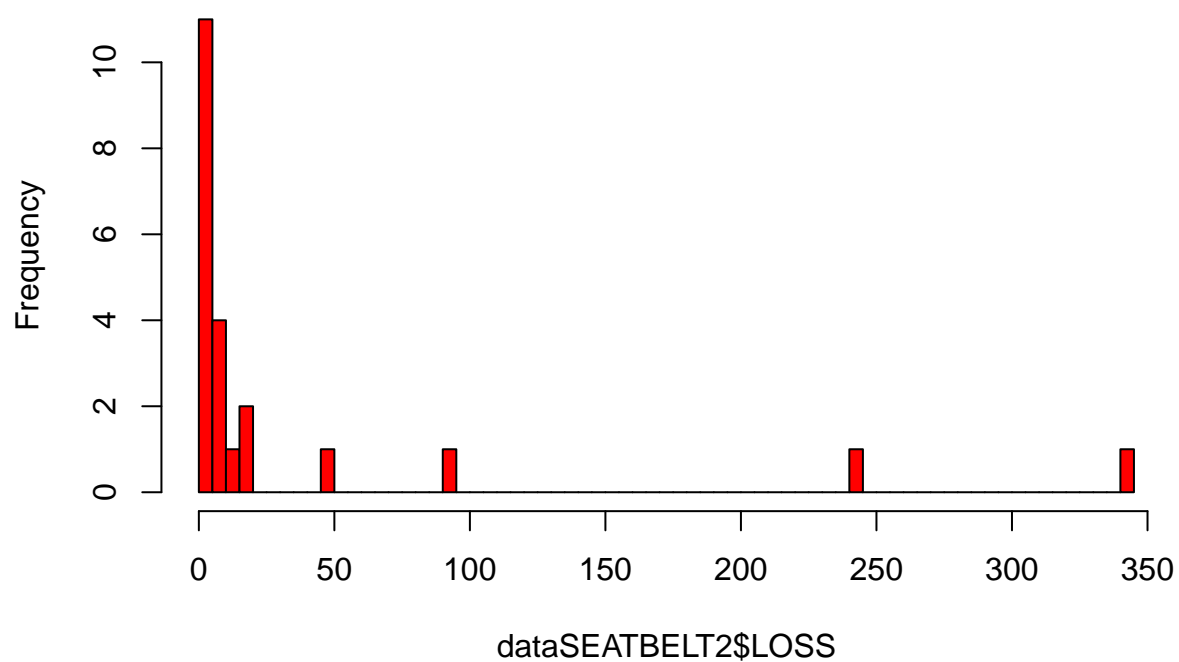
```r
p1=hist(dataSEATBELT1$LOSS,breaks=100,col="blue")
```
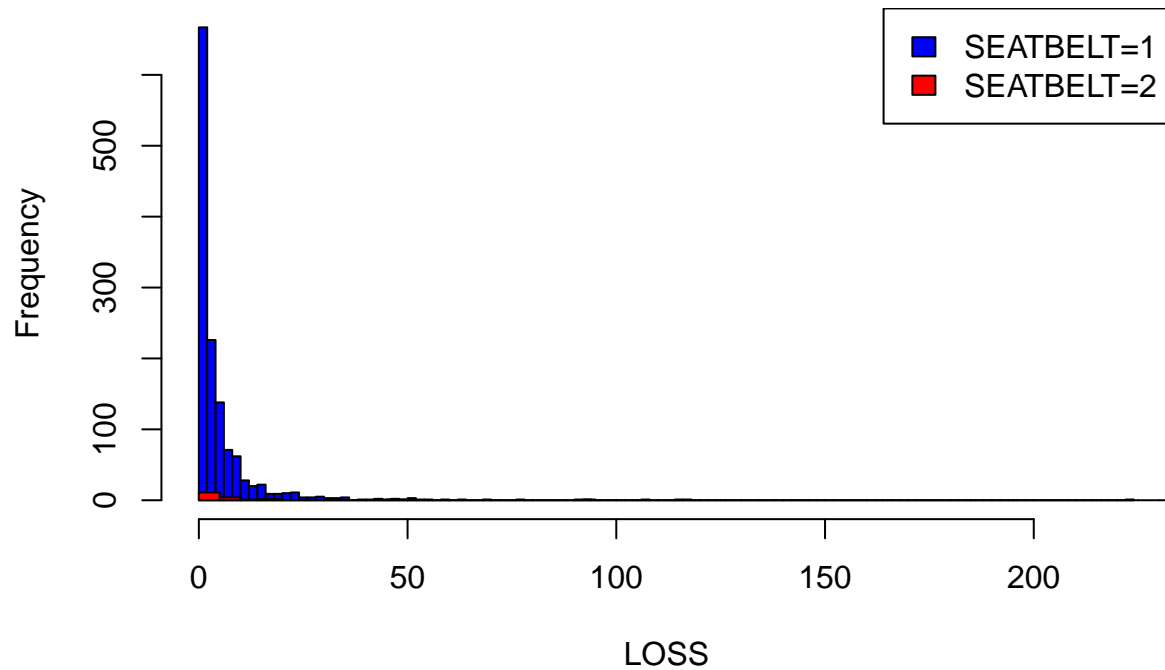
**Histogram of dataSEATBELT1$LOSS**
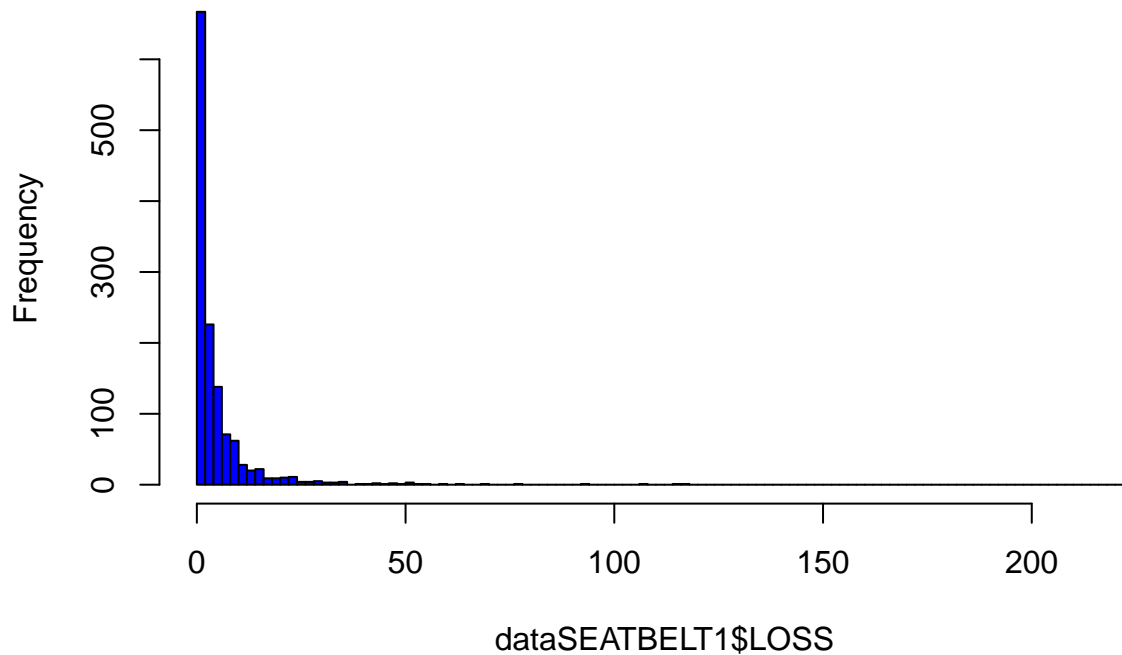


```r
p2=hist(dataSEATBELT2$LOSS,breaks=100,col="red")
```

# Histogram of dataSEATBELT2$LOSS



```
plot(p1,col="blue",xlab="LOSS",main="Histogram of LOSS with different SEATBELT")
plot(p2,col="red",add=T)
legend("topright", c("SEATBELT=1", "SEATBELT=2"), fill=c("blue", "red"))
```

## Histogram of LOSS with different SEATBELT



From the plot we can see these two loss distribution have the same distribution with very likely to be exponential distribution again. However, although the frequency of loss belongs to SEATBELT=2 is very low compared to the loss with SEATBELT=1 one, in this case the loss with SEATBELT=2 one seems having larger tail than the loss with SEATBELT=1 one since most of the extreme value of loss belongs to SEATBELT=1 one.