

## SCS 2111 – Statistical Methods using R

### Lab Session 3 – Answers

```
> # first bullet
> id<-seq(101,150)
>
> # second bullet
> gender<-sample(c("Male","Female"),50,replace=T)
>
> # third bullet
> department<-sample(c("Sales","HR","Accounts"),50,prob=c(0.5,0.25,0.25),
replace=T)
> data<-data.frame(cbind(id,gender,department))
> # fourth bullet
> salary<-numeric(length(id))
> for (i in 1:length(id))
+ {
+ if (department[i]=="Sales") salary[i]<-rnorm(1,15000,1250)
+ if (department[i]=="Accounts") salary[i]<-runif(1,15000,20000)
+ else salary[i]<-rnorm(1,25000,250)
+ }
> data$salary<-round(salary,0)
```

### Part I

```
> head(data)
  id gender department salary
1 101 Female      Sales  25000
2 102 Female        HR  24680
3 103 Female      Sales  24876
4 104 Female      Sales  25242
5 105   Male      Sales  25207
6 106   Male      Sales  24935
```

```
> tail(data)

      id gender department salary
45 145   Male   Accounts  16543
46 146   Male     Sales  25107
47 147 Female     Sales  24538
48 148 Female        HR  25098
49 149 Female   Accounts  16645
50 150   Male   Accounts  19772
```

## Part 2:

```
> data.m<-data[gender=="Male",]
> data.f<-data[gender=="Female",]
> summary(data.m[,4])

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
16540  23350   24930   23350  25110   25420

> summary(data.f[,4])

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
16220  20330   24850   23030  25060   25590

> hist(data.m[,4])
> hist(data.f[,4])
```

Can comment on the values using the summary statistics and the shape of the distribution using the histograms.

## Part 3 :

Same as in part 2.

Note : They can use other appropriate plots also. And please note that the data set can be different for different students.