STAT 40001/MA 59800 Statistical Computing/ Computational Statistics Fall 2013
Homework 3-Solution

Name:
Due : September 26, 2013 PUID:

Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw output). Some of the questions below can be answered with very little or no programming. However, write code that outputs the final answer and does not require any additional paper calculations.

Q.N. 1) Results from an experiment to compare yields (as measured by dried weight of plants) obtained under a control and two different treatment conditions is provided in the data frame `PlantGrowth` in the R dataset.
a) How many observations are recorded in the data set?
b) What is the mean of each of the control and treatment conditions?
c) Create side-by-side boxplots to compare the yields.
Solution:
a) Note that the data set `PlantGrowth` *in the R dataset so we can simply use the R code below to read the data and identify the number of observations*

```
> data(PlantGrowth)
> dim(PlantGrowth)
[1] 30  2
```

Therefore, there are 30 observations measured in two variables.
b) We can use the R code below to find the mean of each gropes:

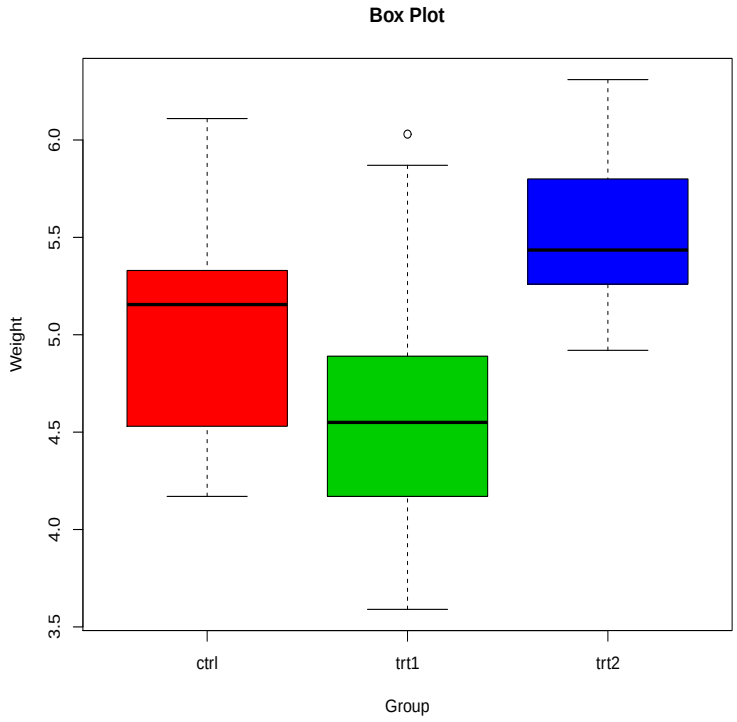
```
> tapply(PlantGrowth$weight,PlantGrowth$group,mean)
 ctrl  trt1  trt2
5.032 4.661 5.526
```

OR

```
> x1=subset(PlantGrowth, group=="ctrl")
> x2=subset(PlantGrowth, group=="trt1")
> x3=subset(PlantGrowth, group=="trt2")
> mean(x1$weight)
[1] 5.032
> mean(x2$weight)
[1] 4.661
> mean(x3$weight)
[1] 5.526
```

c) We can Use the R code below to draw the side-side boxplot as shown below

```
> attach(PlantGrowth)
> names(PlantGrowth)
[1] "weight" "group"
> group<-factor(group)
> plot(weight, group)
> boxplot(weight~group, col=c(2,3,4), ylab="Weight", xlab="Group", main="Box Plot")
```



Q.N. 2) Load `Cars93` from the `MASS` package.
(a) Create density histograms for the variables `Min.Price`, `Max.Price`, `Weight` and `Length` variables using different color for each histogram
(b) Superimpose estimated density curves over the histograms.

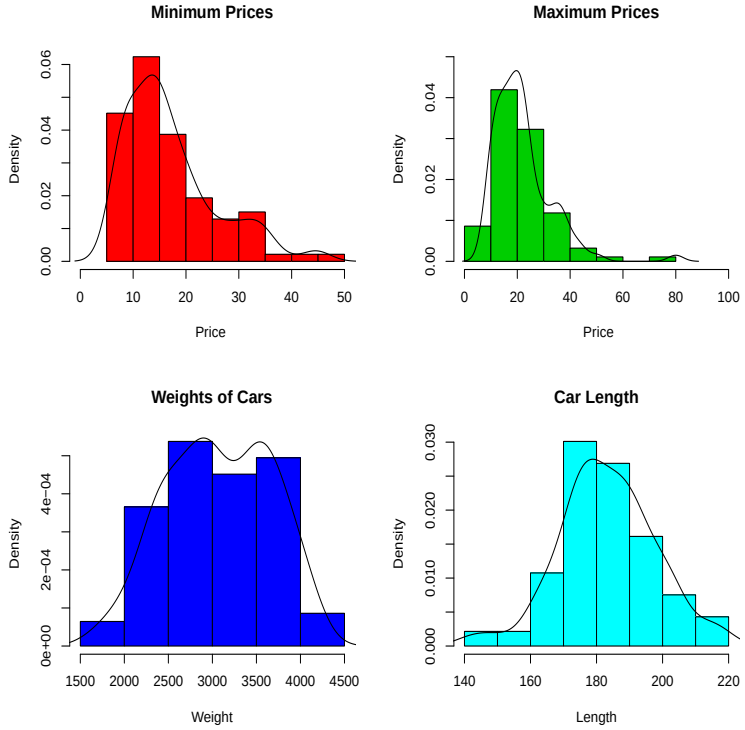
Solution: We can use the code below to load the Cars93 data and find the number of variable being included in the dataset

```
> library(MASS)
> attach(Cars93)
> dim(Cars93)
[1] 93 27
> names(Cars93)
 [1] "Manufacturer"       "Model"              "Type"
 [4] "Min.Price"          "Price"              "Max.Price"
 [7] "MPG.city"           "MPG.highway"        "AirBags"
[10] "DriveTrain"         "Cylinders"          "EngineSize"
[13] "Horsepower"         "RPM"                "Rev.per.mile"
[16] "Man.trans.avail"    "Fuel.tank.capacity" "Passengers"
[19] "Length"             "Wheelbase"          "Width"
[22] "Turn.circle"        "Rear.seat.room"     "Luggage.room"
[25] "Weight"             "Origin"             "Make"
```

It appears that there are 27 variables.

In order to obtain the histograms and density plots displayed below we can use the R code below

```
library(MASS)
attach(Cars93)
par(mfrow=c(2,2))
hist(Min.Price, prob=T, col=2,main="Minimum Prices",xlim=c(0,50),xlab="Price")
lines(density(Min.Price),xlim=c(0,50))
hist(Max.Price,prob=T, col=3,main="Maximum Prices",xlim=c(0,100), ylim=c(0,0.05),
xlab=" Price")
lines(density(Max.Price),xlim=c(0,100))
hist(Weight,prob=T,col=4,main="Weights of Cars",xlim=c(1500,4500),xlab="Weight")
lines(density(Weight))
hist(Length,prob=T,col=5,main="Car Length",xlab="Length")
lines(density(Length))
```



Q.N. 3) Generate 100 random numbers from normal distribution with mean 100 and standard deviation 10. How many are 2 standard deviations from the mean (smaller than 80 or bigger than 120)?
Solution: We can use the code below in R to generate 100 random numbers from normal random variable with mean 100 and standard deviation 10 and count the number of observations in the given range

```
> set.seed(123)
> x=rnorm(100,mean=100,sd=10)
> sum(x<80 | x>120)
[1] 4
```

Q.N. 4) If Z standard normal distribution find the following
a) $P(Z \leq 2.2)$
b) $P(-1 \leq Z \leq 1)$
c) $P(-1 < Z < 1)$
d) $P(Z > 2.5)$
e) b such that $P(-b \leq Z \leq b) = 0.90$
Solution: We can use the R code below to answer each question:

```
#a)
> pnorm(2.2)
[1] 0.9860966
#b)
> pnorm(1)-pnorm(-1)
[1] 0.6826895
#c)
> pnorm(1)-pnorm(-1)
[1] 0.6826895
#d)
> 1-pnorm(2.5)
[1] 0.006209665
#e)
> qnorm(.95)
[1] 1.644854
```

Q.N.5) The `abd` package in R contains data sets related to Biological studies. The data frame `TwoKids` in the `abd` package has of the information about the number of boys in two-child families. Display the information contain in the data choosing appropriate graphical method.
Solution: After installing the `abd` *package we access the* `TwoKids` *data and display the information graphically using Pie Chart using R code below*

```
library(abd)
data("TwoKids",package="abd")
attach(TwoKids)
names(count)=c("0 boys","1 boy","2 boys")
pie(count,col=c(1,2,3),main="Number of Boys in Two-Child Families")
```