EXP 6:

```
@RELATION iris
@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

The Data of the ARFF file looks like the following:

```
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
```

EXP 7:

```
Dataset student .arff
@relation student
@attribute age {<30,30-40,>40}
@attribute income {low, medium,
high}
@attribute student {yes, no}
@attribute credit-rating {fair,
excellent}
@attribute buyspc {yes, no}
@data
%
<30, high, no, fair, no
<30, high, no, excellent, no
30-40, high, no, fair, yes
>40, medium, no, fair, yes
>40, low, yes, fair, yes
>40, low, yes, excellent, no
30-40, low, yes, excellent, yes
<30, medium, no, fair, no
<30, low, yes, fair, no
>40, medium, yes, fair, yes
<30, medium, yes, excellent, yes
30-40, medium, no, excellent,
yes
```

```
30-40, high, yes, fair, yes
>40, medium, no, excellent, no
%
```

EXP 3:

```
Data set employee.arff:
@relation employee
@attribute age {25, 27, 28, 29,
30, 35, 48}
@attribute
salary{10k,15k,17k,20k,25k,30k,3
5k,32k}
@attribute performance {good,
avg, poor}
@data
%
25, 10k, poor
27, 15k, poor
27, 17k, poor
28, 17k, poor
29, 20k, avg
30, 25k, avg
29, 25k, avg
30, 20k, avg
35, 32k, good
48, 35k, good
```

```
48, 32k,good
%
```

WEEK 8:

```
b)
input <-
mtcars[,c("am","mpg","hp")]
print(head(input))

input <- mtcars

result <- aov(mpg~hp*am,data =
input)
print(summary(result))

result <- aov(mpg~hp+am,data =
input)
print(summary(result))

result1 <- aov(mpg~hp*am,data =
input)
result2 <- aov(mpg~hp+am,data =
input)
print(anova(result1,result2))
```

```
result <- aov(mpg~hp*am,data =
input)
> print(summary(result))
o/p:

Df Sum Sq Mean Sq F value
Pr(>F)
hp
am
hp:am
1  678.4   678.4  77.391 1.50e-
09 ***
1  202.2   202.2  23.072 4.75e-
05 ***
1    0.0     0.0   0.001
0.981
Residuals   28  245.4     8.8
---
Signif. codes:  0 '***' 0.001
'**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
result <- aov(mpg~hp+am,data =
input)
> print(summary(result))
Df Sum Sq Mean Sq F value
Pr(>F)
```

```
hp
am
1  678.4   678.4   80.15 7.63e-
10 ***
1  202.2   202.2   23.89 3.46e-
05 ***
Residuals   29  245.4     8.5
---
Signif. codes:  0 '***' 0.001
'**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
result1 <- aov(mpg~hp*am,data =
input)
> result2 <- aov(mpg~hp+am,data
= input)
> # Compare the two models.
> print(anova(result1,result2))
Analysis of Variance Table
Model 1: mpg ~ hp * am
Model 2: mpg ~ hp + am
Res.Df    RSS Df  Sum of Sq
F Pr(>F)
1     28 245.43
2     29 245.44 -1 -0.0052515
6e-04 0.9806
```

WEEK 9:

d) Poisson Regression .

Source code:

```
input <- warpbreaks
print(head(input))
output <-glm(formula = breaks ~
wool+tension, data = warpbreaks,
family = poisson)
print(summary(output))
```

output:

```
input <- warpbreaks
> print(head(input))
  breaks wool tension
1     26    A       L
2     30    A       L
3     54    A       L
4     25    A       L
5     70    A       L
6     52    A       L
> output <-glm(formula = breaks
~ wool+tension, data =
warpbreaks,
+                family = poisson)
> print(summary(output))
```

```
Call:
glm(formula = breaks ~ wool +
tension, family = poisson, data
= warpbreaks)

Deviance Residuals:
    Min        1Q   Median
3Q       Max   -3.6871  -1.6503
-0.4269   1.1902   4.2616

Coefficients:
            Estimate Std. Error
z value Pr(>|z|)
(Intercept)  3.69196    0.04541
81.302  < 2e-16 ***
woolB       -0.20599    0.05157
-3.994 6.49e-05 ***
tensionM    -0.32132    0.06027
-5.332 9.73e-08 ***
tensionH    -0.51849    0.06396
-8.107 5.21e-16 *** ---
Signif. codes:  0 '***' 0.001
'**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for poisson family taken to be 1)
WEEK 10.a)

Source code:

```
snowfall <-
c(790,1170.8,860.1,1330.6,630.4,
911.5,683.5,996.6,783.2,982,881.
8,1021)
snowfall_timeseries<-
ts(snowfall,start =
c(2013,1),frequency = 12)
print(snowfall_timeseries)
png(file = "snowfall.png")
  plot(snowfall_timeseries)
dev.off()
```

o/p:

```
> print(snowfall_timeseries)
Jan
Feb    Mar    Apr    May    Jun
Jul    Aug    Sep    Oct
2013  790.0 1170.8  860.1 1330.6
630.4  911.5  683.5  996.6
783.2  982.0
```

10.B)

```
xvalues <-
c(1.6,2.1,2,2.23,3.71,3.25,3.4,3
.86,1.19,2.21)
yvalues <-
c(5.19,7.43,6.94,8.11,18.75,14.8
8,16.06,19.12,3.21,7.58)
png(file = "nls.png")
plot(xvalues,yvalues)
model <- nls(yvalues ~
b1*xvalues^2+b2,start = list(b1
= 1,b2 = 3))
new.data <- data.frame(xvalues =
seq(min(xvalues),max(xvalues),le
n = 100))
lines(new.data$xvalues,predict(m
odel,newdata = new.data))
dev.off()
print(sum(resid(model)^2))
print(confint(model))
10.c)
data("iris")
install.packages("caret")
install.packages("C50")
library(caret)
```

```
library(C50)
set.seed(7)
inTraininglocal<-
createDataPartition(iris$Species
,p=.70,list = F)
training<-iris[inTraininglocal,]
testing<-iris[-inTraininglocal,]
model<-C5.0(Species~.,data =
training)
summary(model)
pred<-
predict.C5.0(model,testing[,-5])
#type ="prob"
a<-table(testing$Species,pred)
sum(diag(a))/sum(a)
plot(model)
dev.off()
```

11)a) Normal Distribution

```
dnorm
x <- seq(-10, 10, by = .1)
y <- dnorm(x, mean = 2.5, sd =
0.5)
png(file = "dnorm.png")
plot(x,y)
dev.off()
```

Pnorm

```
x <- seq(-10,10,by = .2)
y <- pnorm(x, mean = 2.5, sd =
2)
png(file = "pnorm.png")
plot(x,y)
dev.off()
```

qnorm

```
x <- seq(0, 1, by = 0.02)
y <- qnorm(x, mean = 2, sd = 1)
png(file = "qnorm.png")
plot(x,y)
dev.off()
```

rnorm

```
y <- rnorm(50)
png(file = "rnorm.png")
hist(y, main = "Normal
DIstribution");
dev.off()
```

b) Binomial Distribution

dbinom

```
x <- seq(0,50,by = 1)
```

```
y <- dbinom(x,50,0.5)
png(file = "dbinom.png")
plot(x,y)
dev.off()
pbinom:
x <- pbinom(26,51,0.5)
print(x)
output:
print(x)
[1] 0.610116
qbinom:
x <- qbinom(0.25,51,1/2)
print(x)
output:
print(x)
[1] 23
rbinom:
x <- rbinom(8,150,.4)
print(x)
output:
print(x)
[1] 68 59 55 49 51 59 53 55
```

12)a) χ2-test

```
library("MASS")
print(str(Cars93))
car_data<-
data.frame(Cars93$AirBags,
Cars93$Type)
car_data = table(Cars93$AirBags,
Cars93$Type)
print(car_data)
print(chisq.test(car_data))
b) t-test
x  <- c(0.593, 0.142, 0.329,
0.691, 0.231, 0.793, 0.519,
0.392, 0.418)
t.test(x, alternative="greater",
mu=0.3)
output:
One Sample t-test
data:  x
t = 2.2051, df = 8, p-value =
0.02927
alternative hypothesis: true
mean is greater than 0.3
```

```
95 percent confidence interval:
0.3245133
Inf
sample estimates:
mean of x
0.4564444
F Test:
install.packages("randomForest")
library(party)
print(head(readingSkills))
library(party)
library(randomForest)
output.forest <-
randomForest(nativeSpeaker ~ age
+ shoeSize + score,
data = readingSkills)
print(output.forest)



output:
print(output.forest)
Call:
```

```
randomForest(formula =
nativeSpeaker ~ age + shoeSize +
score,
Type of random forest:
classification
Number of trees: 500
No. of variables tried at each
split: 1
OOB estimate of  error rate:
1.5%
Confusion matrix:
no yes class.error
no  99   1        0.01
yes  2  98
0.02
```