Factors

```
gender <- c(rep("male",20), rep("female", 30))
gender <- factor(gender)
print(gender)
plot(gender)
```

LDA(Discriminant Analysis)

```
library(MASS)
url <- 'http://www.biz.uiowa.edu/faculty/jledolter/DataMining/admission.csv'
admit <- read.csv(url)
print(head(admit))
adm=data.frame(admit)
plot(adm$GPA,adm$GMAT,col=c("red","green","black"))
m1=lda(De~.,adm)
print(m1)
m2=qda(De~.,adm)
print(m2)
library(klaR)
partimat(De~.,data=adm,method="lda")
```

Another example

```
credit <- read.csv("http://www.biz.uiowa.edu/faculty/jledolter/DataMining/germancredit.csv")
print(head(credit,2)) # See details about codification in the attached documentation.
cred1=credit[, c("Default","duration","amount","installment","age")]
print(head(cred1))
print(summary(cred1))
hist(cred1$duration)
hist(cred1$amount)
hist(cred1$installment)
hist(cred1$age)
```

Anova

What is ANOVA?

Analysis of Variance (ANOVA) is a statistical technique, commonly used to studying differences between two or more group means. ANOVA test is centred on the different sources of variation in a typical variable. ANOVA in R primarily provides evidence of the existence of the mean equality between the groups. This statistical method is an extension of the t-test. It is used in a situation where the factor variable has more than one group.

Program

```
group1<-c(2,3,7,3,6)

group2<-c(10,20,4,10,5)

group3<-c(10,13,14,13,15)

combined_groups<-data.frame(cbind(group1,group2,group3))

print(combined_groups)

print(summary(combined_groups))

stact_group<-stack(combined_groups)

print(stact_group)

anova_res<-aov(values~ind,data=stact_group)

print(anova_res)

print(summary(anova_res))
```

Correlation

Correlation analysis is a statistical method used to evaluate the strength of relationship between two quantitative variables. A high correlation means that two or more variables have a strong relationship with each other, while a weak correlation means that the variables are hardly related. In other words, it is the process of studying the strength of that relationship with available statistical data.

Program

```
Lungcapdata<-read.table("D:/LungCapData.txt",header = T)

print(Lungcapdata)

plot(Lungcapdata$Age,Lungcapdata$LungCap,main="Scatterplot",las=1)

print(cor(Lungcapdata$Age,Lungcapdata$LungCap,method="pearson"))

print(cor(Lungcapdata$Age,Lungcapdata$LungCap,method="spearman"))

print(cor(Lungcapdata$Age,Lungcapdata$LungCap,method="kendall"))

print(cor.test(Lungcapdata$Age,Lungcapdata$LungCap,method="spearman",exact=F))
```

Regression

Regression analysis is a very widely used statistical tool to establish a relationship model between two variables. One of these variable is called predictor variable whose value is gathered through experiments. The other variable is called response variable whose value is derived from the predictor variable.

Program

```
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)

y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

# Apply the lm() function.

relation <- lm(y~x)

print(relation)

print(summary(relation))

a <- data.frame(x = 170)

result <-  predict(relation,a)

print(result)

plot(y,x,col = "blue",main = "Height & Weight Regression",

    abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg",ylab = "Height in cm")
```