```
#Simulation
set.seed(10101)
n<-1000
p<-10
Sig<-rWishart(1,p+1,diag(1,p))[,,1]
library(MASS)
X<-mvrnorm(n,mu=rep(0,p),Sigma=Sig)
pairs(X,pch=19,cex=0.5)
beta<-rep(0,p+1)
beta[1]<-10
beta[4:7]<-c(3,-2,1,2.5)
Xmat<-cbind(rep(1,n),X)

sig<-2.0

mean.Y<-Xmat %*% beta

Y<-rnorm(n)*sig+mean.Y

#####################################################
#Exploration

par(mfrow=c(4,3),mar=c(2, 4, 2, 2) + 0.1)
for(j in 1:p){
str<-substitute(X[index], list(index=j))
plot(X[,j],Y,pch=19,cex=0.8,xlab=str)
title(substitute("Y vs "*X[index]*" ",list(index=j)))
}
par(mfrow=c(1,1),mar=c(5, 4, 4, 2) + 0.1)

cor(cbind(Y,X))

#####################################################
#Fitting

mod0<-lm(Y~1)
summary(mod0)

mod1<-lm(Y~X)
summary(mod1)

names(mod1)
names(summary(mod1))

coef(mod1)
coef(summary(mod1))
summary(mod1)[[4]]

resid(mod1)

anova(mod1)

drop1(mod1,test="F")

#####################################
my.data<-data.frame(X,Y)

mod1<-lm(Y~.,data=my.data)
summary(mod1)
mod2<-lm(Y~X1,data=my.data)
summary(mod2)


mod3<-lm(Y~X1+X2,data=my.data)
summary(mod3)
drop1(mod3,test="F")
anova(mod3)

mod3a<-lm(Y~-1+X1+X2,data=my.data)
```

```
summary(mod3a)
drop1(mod3a,test="F")
anova(mod3a)


mod4<-lm(Y~X1+X2+X1:X2,data=my.data)
summary(mod4)
drop1(mod4,test="F")
anova(mod4)

mod5<-lm(Y~X1+X2+X2^2,data=my.data)
summary(mod5)
mod5<-lm(Y~X1+X2+I(X2^2),data=my.data)
summary(mod5)
drop1(mod5,test="F")
anova(mod5)

mod6<-lm(Y~(X1+X2)^2,data=my.data)
summary(mod6)

mod7<-lm(Y~(X1+X2+X3)^2,data=my.data)
summary(mod7)
drop1(mod7,test="F")
anova(mod7)

mod8<-lm(Y~X1*X2*X3,data=my.data)
summary(mod8)
drop1(mod8,test="F")
anova(mod8)

##########################################
A<-factor(sample(0:1,size=n,rep=T),labels=c("M","F"))
B<-factor(sample(1:5,size=n,rep=T),labels=c("TMT1","TMT2","TMT3","TMT4","TMT5"))

my.data<-cbind.data.frame(my.data,A,B)

mod9<-lm(Y~A+B*X1,data=my.data)
summary(mod9)
drop1(mod9,test="F")

mod10<-lm(Y~A*B,data=my.data)
summary(mod10)

anova(mod10)
drop1(mod10,test="F")

##########################################

mod11.form<-formula(Y~X+I(X^2))
summary(lm(mod11.form,data=my.data))

fstr<-paste("Y~",paste("X",1:10,sep="",collapse="+"))
fstr.form<-as.formula(fstr)
summary(lm(fstr.form,data=my.data))

fstr.new<-paste(fstr,"+",paste("I(X",1:10,"^2",")",sep="",collapse="+"),sep="")
fstr.form.new<-as.formula(fstr.new)

mod12<-lm(fstr.form.new,data=my.data)
summary(mod12)

##########################################
step(mod12)
step(mod12,k=log(n))

install.packages("leaps")
library(leaps)
subsets <- regsubsets(fstr.form.new, data=my.data)
plot(subsets)
```

```
################################################

mod.final<-step(mod12,k=log(n))
summary(mod.final)

plot(mod.final)
par(mfrow=c(2,2))
plot(mod.final)
par(mfrow=c(1,1))

Yfit<-fitted(mod.final)
plot(Y,Yfit,pch=19,cex=0.8,ylab=expression(hat(Y)))
abline(0,1,col="red")

res.final<-residuals(mod.final)
hist(res.final)

st.res.final<-rstandard(mod.final)
hist(st.res.final,xlab="Standardized Residual",main="Standardized Res.")

plot(Yfit,st.res.final,pch=19,cex=0.8,xlab=expression(hat(Y)),ylab="St.resid.")

X.final<-model.matrix(mod.final)
Xselected<-c(3,4,5,6)
par(mfrow=c(2,2),mar=c(2, 4, 2, 2) + 0.1)
for(j in 2:ncol(X.final)){
jvar<-Xselected[j-1]
str<-substitute(X[index], list(index=jvar))
plot(X.final[,j],st.res.final,pch=19,cex=0.8,xlab=str,ylab="St.resid.")
title(substitute("St. Resid. vs "*X[index]*" ",list(index=jvar)))
abline(h=c(-2,2),lty=2,col="red")
}
par(mfrow=c(1,1),mar=c(5, 4, 4, 2) + 0.1)

################################################
X.means<-apply(X.final[,3:5],2,mean)
new.X3<-seq(-7.5,7.5,length=1001)
new.data<-data.frame(X3=new.X3,X4=rep(X.means[1],1001),
                     X5=rep(X.means[2],1001),X6=rep(X.means[3],1001))
Ypred<-predict(mod.final,newdata=new.data)
plot(new.X3,Ypred,xlab=expression(X[3]),ylab=expression(hat(Y)),type="l")

xl<-expression(X[3])
yl<-expression(hat(Y))
Ypred<-predict(mod.final,newdata=new.data,interval="confidence")
plot(new.X3,Ypred[,1],xlab=xl,ylab=yl,type="l")
lines(new.X3,Ypred[,2],xlab=xl,ylab=yl,type="l",col="red",lty=2)
lines(new.X3,Ypred[,3],xlab=xl,ylab=yl,type="l",col="red",lty=2)
title("Confidence interval for fitted observation")

Ypred<-predict(mod.final,newdata=new.data,interval="prediction")
plot(new.X3,Ypred[,1],xlab=xl,ylab=yl,type="l")
lines(new.X3,Ypred[,2],xlab=xl,ylab=yl,type="l",col="red",lty=2)
lines(new.X3,Ypred[,3],xlab=xl,ylab=yl,type="l",col="red",lty=2)
title("Prediction interval for fitted observation")
```