Multiple Regression

- **First-Order Model with Two Predictor Variables.**

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i. \qquad i = 1, 2, \ldots, n.$$

where:

1. Y_i is the value of the response variable in the ith trial.
2. β_0, β_1, and β_2 are parameters.
3. X_{i1} is the value of the first predictor variable in the ith trial.
4. X_{i2} is the value of the second predictor variable in the ith trial.
5. ϵ_i is a random error term with $E\{\epsilon_i\} = 0$ and variance σ^2.
6. The random errors ϵ_i are uncorrelated. That is, $Cov(\epsilon_i, \epsilon_j) = 0$

- **Dwaine Studios Example.** Dwaine Studios, Inc, operates portrait studios in 21 cities of medium size. These studios specialize in portraits of children. The company is considering an expansion into other cities of medium size and wishes to investigate whether sales (Y) in a community can be predicted from the number of persons aged 16 or younger in the community (X_1) and the per capita disposable personal income in the community (X_2). The data are stored in 'DwaineStudios.csv' file.

```
data=read.csv("Dwaine.csv",header=T)      # Dwaine Studios example on page 236
attach(data)
y=sales           # Sales of portraits of children in a community
x1=young          # Number of persons aged 16 or younger in the community
x2=income         # Per capita disposable personal income in the community
n=length(sales)
pairs(cbind(x1,x2,y))   # Creates a scatter plot matrix
pairs(data)
cor(data) # Creates a correlation matrix

X=matrix(c(rep(1,n),x1,x2),ncol=3) # The design matrix X
t(X)%*%X
B=solve(t(X)%*%X)%*%(t(X)%*%y) # Estimates of the regression coefficientspair

#Or we can use
results=lm(y~x1+x2)
summary(results)

# Diagnostics
par(mfrow=c(2,2))
plot(results$fitted,results$residuals)
plot(x1,results$residuals)
plot(x2,results$residuals)
plot(x1*x2,results$residuals) # Checking for interaction effects

plot(results$fitted,abs(results$residuals)) # Checking constant variance
qqnorm(results$residuals) # Checking normality
shapiro.test(results$residuals)

anova(results)
# F Test for Regression Relation
# Coefficient of multiple determination
# Adjusted coefficient of multiple determination
# Coefficient of multiple correlation

#Prediction
new=data.frame(x1=65.4,x2=17.6)
predict(results,new)
predict(results,newdata=new,interval="confidence")
predict(results,newdata=new,interval="prediction")

new2=data.frame(x1=c(65.4,53.1),x2=c(17.6,17.7))
predict(results,newdata=new2,interval="prediction",level=(1-.10/2)) # Bonferroni Intervals
```