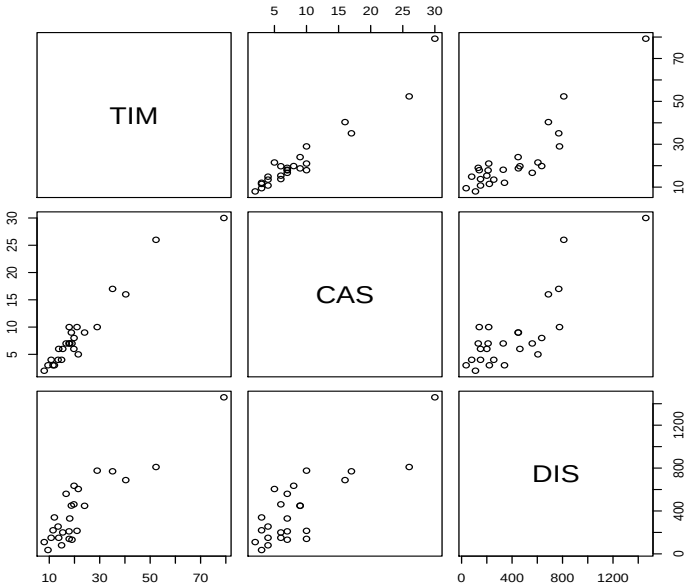
Some matrix operations in R

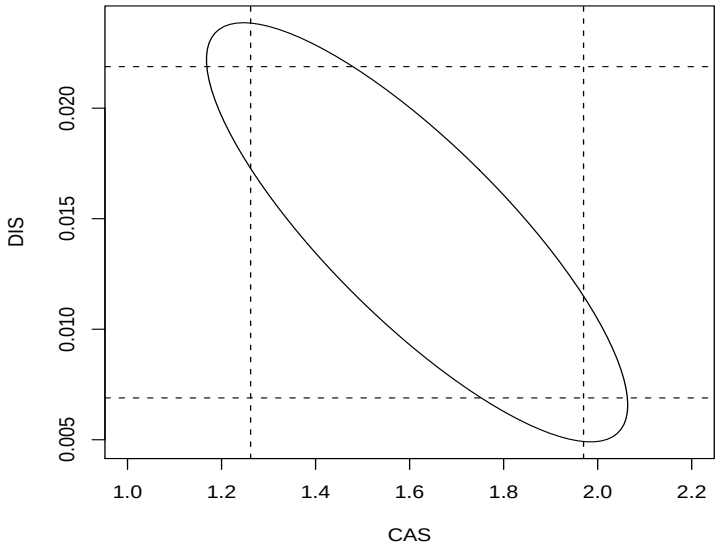
```
> #read in the data from a csv file
> dat1=read.table("Data Sets/Chapter 3/Examples/ex-3-1.csv", header=T, sep=",")
> str(dat1)  #checking data format

'data.frame':        25 obs. of  4 variables:
 $ Observation        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Delivery.Time..y   : num  16.7 11.5 12.0 14.9 13.8 ...
 $ Number.of.Cases..x1: int  7 3 3 4 6 7 2 7 30 5 ...
 $ Distance..x2..ft.  : int  560 220 340 80 150 330 110 210 1460 605 ...

> TIM=dat1[,2] #assign the second column in dat1 to be variable TIM
> CAS=dat1[,3] #assign the third column in dat1 to be variable CAS
> DIS=dat1[,4] #assign the forth column to be variable DIS
> EX3=data.frame(TIM, CAS, DIS)
> head(EX3)  #printout the first few row in dat1

    TIM CAS DIS
1 16.68   7 560
2 11.50   3 220
3 12.03   3 340
4 14.88   4  80
5 13.75   6 150
6 18.11   7 330

> pairs(EX3)
```



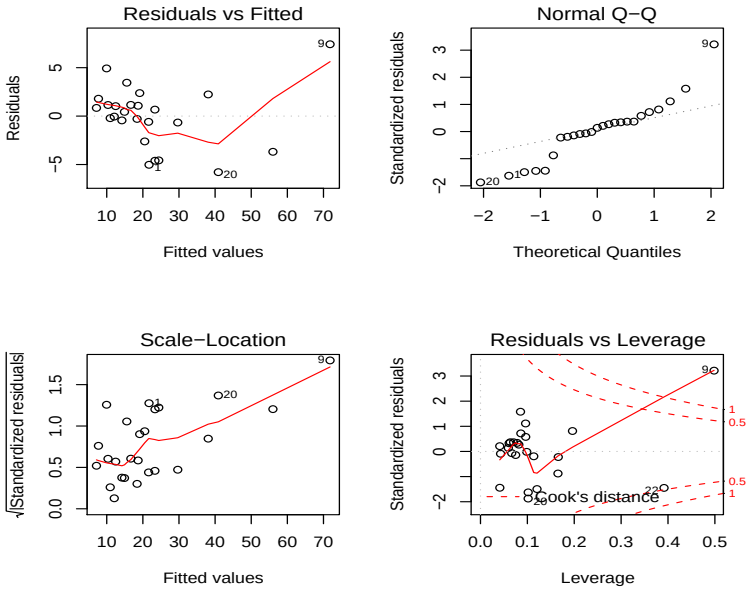
```
> fm1=lm(TIM~CAS+DIS, data=EX3)
> anova(fm1)

Analysis of Variance Table

Response: TIM
          Df Sum Sq Mean Sq F value    Pr(>F)
CAS        1 5382.4  5382.4 506.619 < 2.2e-16
DIS        1  168.4   168.4  15.851 0.0006312
Residuals 22  233.7    10.6

> summary(fm1)

Call:
lm(formula = TIM ~ CAS + DIS, data = EX3)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7880 -0.6629  0.4364  1.1566  7.4197

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.341231   1.096730   2.135 0.044170
CAS         1.615907   0.170735   9.464 3.25e-09
DIS         0.014385   0.003613   3.981 0.000631

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared: 0.9596,        Adjusted R-squared: 0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16

> fitted(fm1)        #Y.hat

        1         2         3         4         5         6         7
21.708084 10.353615 12.079794  9.955646 14.194398 18.399574  7.155376
        8         9        10        11        12        13        14
16.673395 71.820294 19.123587 38.092507 21.593041 12.472991 18.682464
       15        16        17        18        19        20        21
23.328798 29.662928 14.913640 15.551379  7.706807 40.887970 20.514179
       22        23        24        25
56.006528 23.357568 24.402854 10.962584

> resid(fm1)        #e :residual

         1          2          3          4          5          6
-5.0280843  1.1463854 -0.0497937  4.9243539 -0.4443983 -0.2895743
         7          8          9         10         11         12
 0.8446235  1.1566049  7.4197062  2.3764129  2.2374930 -0.5930409
        13         14         15         16         17         18
 1.0270093  1.0675359  0.6712018 -0.6629284  0.4363603  3.4486213
```

```
        19         20         21         22         23         24
 1.7931935 -5.7879699 -2.6141789 -3.6865279 -4.6075679 -4.5728535
        25
-0.2125839

> df.residual(fm1)  #n-p

[1] 22

> coef(fm1)            #beta.hat's

(Intercept)         CAS         DIS
 2.34123115  1.61590721  0.01438483

> SSres=sum(resid(fm1)^2);SSres        #sum of squares due to residuals

[1] 233.7317

> MSres=SSres/df.residual(fm1); MSres  #Residual Mean Square

[1] 10.62417

> sqrt(MSres)

[1] 3.259473

> X=model.matrix(fm1) #assign the model matrix of the linear model to X
> head(X)

  (Intercept) CAS DIS
1           1   7 560
2           1   3 220
3           1   3 340
4           1   4  80
5           1   6 150
6           1   7 330

> t(X) %*% X        #X'X

            (Intercept)    CAS     DIS
(Intercept)          25    219   10232
CAS                 219   3055  133899
DIS               10232 133899 6725688

> t(X) %*% TIM         #X'Y

                 [,1]
(Intercept)    559.60
CAS           7375.44
DIS         337071.69
```

```
> solve(t(X) %*% X) %*% t(X) %*% TIM   #inv(X'X)X'Y

                 [,1]
(Intercept) 2.34123115
CAS         1.61590721
DIS         0.01438483

> solve(t(X) %*% X)     #inv(X'X)

              (Intercept)           CAS           DIS
(Intercept)  1.132152e-01 -4.448593e-03 -8.367257e-05
CAS         -4.448593e-03  2.743783e-03 -4.785709e-05
DIS         -8.367257e-05 -4.785709e-05  1.228745e-06

> C=solve(t(X) %*% X)   #assign C to be the inverse of the X'X matrix
> sqrt(MSres*diag(C))   #Standard error of the estimators of beta's

(Intercept)         CAS         DIS
1.096730168 0.170734918 0.003613086

>                       #diag(C) diagonal elemenets of matrix C
>
> vcov(fm1)  #variance and covariance matrix for beta.hat's

              (Intercept)           CAS           DIS
(Intercept)  1.2028170618 -0.0472625981 -8.889514e-04
CAS         -0.0472625981  0.0291504123 -5.084417e-04
DIS         -0.0008889514 -0.0005084417  1.305439e-05

> sqrt(diag(vcov(fm1))) # standard errors of beta.hat's

(Intercept)         CAS         DIS
1.096730168 0.170734918 0.003613086

> confint(fm1)   #confident interval for the estimators of beta's

                 2.5 %     97.5 %
(Intercept) 0.066751987 4.61571030
CAS         1.261824662 1.96998976
DIS         0.006891745 0.02187791

> #install.packages("ellipse")  #install the package "ellipse"
> library(ellipse)  #call-in the library
> plot(ellipse(fm1, which=c(2,3)), type="l", xlim=c(1,2.2))
>                   #confidence region for CAS and DIS
> abline(v=confint(fm1)[2,], lty=2) #confidence interval for CAS
> abline(h=confint(fm1)[3,], lty=2) #confidence interval for DIS
> cor(CAS, DIS)

[1] 0.824215
```




```
> par(mfrow=c(2,2))
> plot(fm1, ask=F)
```




```
> fm2=lm(TIM~CAS)
> anova(fm2)

Analysis of Variance Table

Response: TIM
          Df Sum Sq Mean Sq F value   Pr(>F)
```

```
CAS        1 5382.4  5382.4  307.85 8.22e-15
Residuals 23  402.1    17.5

> anova(fm2, fm1)

Analysis of Variance Table

Model 1: TIM ~ CAS
Model 2: TIM ~ CAS + DIS
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     23 402.13
2     22 233.73  1    168.40 15.851 0.0006312
```