The Analysis of Variance (ANOVA) Table
for an Unbalanced Two Factor Experiment

	Storage Temperature	
Storage Time	**20°C**	**30°C**
3 months	**2 5**	**9 12 15**
6 months	**6 6 7 7**	**16**

```
time=factor(rep(c(3,6),each=5))
temp=factor(rep(c(20,30,20,30),c(2,3,4,1)))
y=c(2,5,9,12,15,6,6,7,7,16)
d=data.frame(time,temp,y)
d
   time temp  y
1     3   20  2
2     3   20  5
3     3   30  9
4     3   30 12
5     3   30 15
6     6   20  6
7     6   20  6
8     6   20  7
9     6   20  7
10    6   30 16

o=lm(y~time+temp+time:temp,data=d)
```

```
anova(o)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value   Pr(>F)
time        1   0.10   0.100  0.0255 0.878292
temp        1 158.42 158.420 40.4477 0.000709 ***
time:temp  1   0.48   0.480  0.1226 0.738243
Residuals  6  23.50   3.917

X1=matrix(1,nrow=10,ncol=1)
X1
      [,1]
 [1,]    1
 [2,]    1
 [3,]    1
 [4,]    1
 [5,]    1
 [6,]    1
 [7,]    1
 [8,]    1
 [9,]    1
[10,]    1
```

```
X2=model.matrix(~time)
X2
   (Intercept) time6
1            1     0
2            1     0
3            1     0
4            1     0
5            1     0
6            1     1
7            1     1
8            1     1
9            1     1
10           1     1
```

```
X3=model.matrix(~time+temp)
X3
   (Intercept) time6 temp30
1            1     0      0
2            1     0      0
3            1     0      1
4            1     0      1
5            1     0      1
6            1     1      0
7            1     1      0
8            1     1      0
9            1     1      0
10           1     1      1
```

```
X4=model.matrix(~time+temp+time:temp)
X4
   (Intercept) time6 temp30 time6:temp30
1            1     0      0            0
2            1     0      0            0
3            1     0      1            0
4            1     0      1            0
5            1     0      1            0
6            1     1      0            0
7            1     1      0            0
8            1     1      0            0
9            1     1      0            0
10           1     1      1            1
```

```
Px=function(X)
{
  X%*%solve(t(X)%*%X)%*%t(X)
}

P1=Px(X1)
P1
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [2,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [3,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [4,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [5,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [6,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [7,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [8,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [9,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
[10,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
```

P2=Px(X2)
P2

	1	2	3	4	5	6	7	8	9	10
1	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0
2	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0
3	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0
4	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0
5	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2
7	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2
8	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2
9	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2
10	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2

P3=Px(X3)

P4=Px(X4)

```
drop(t(y)%*%(P2-P1)%*%y)
[1] 0.1
drop(t(y)%*%(P3-P2)%*%y)
[1] 158.42
drop(t(y)%*%(P4-P3)%*%y)
[1] 0.48

anova(o)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
time       1   0.10   0.100  0.0255 0.878292
temp       1 158.42 158.420 40.4477 0.000709 ***
time:temp  1   0.48   0.480  0.1226 0.738243
Residuals  6  23.50   3.917
```

To understand the null hypothesis that is being tested by the ANOVA table time test, we should examine the noncentrality parameter of that test.

The noncentrality parameter for the test of time is the squared length of the vector P2 E(y) - P1 E(y), divided by sigma^2.

Note that E(y) is of the form

mu_11
mu_11
mu_12
mu_12
mu_12
mu_21
mu_21
mu_21
mu_21
mu_22

Thus, P1 E(y) is the vector with 10 components each identical to

(2*mu_11+3*mu_12+4*mu_21+mu_22)/10.

P2 E(y) is has two distinct components:

(2*mu_11+3*mu_12)/5 repeated five times

and

(4*mu_21+mu_22)/5 repeated five times.

It follows that the noncentrality parameter will be zero if and only if

2*mu_11+3*mu_12=4*mu_21+mu_22.

Thus the time test in the ANOVA table is testing

H_0: 2*mu_11+3*mu_12=4*mu_21+mu_22.

This is not a scientifically interesting test.

```
#Now suppose the data are balanced.
#For balanced data, the sequential sums of squares
#are the same regardless of the order of the
#terms in the anova table.

time=factor(rep(c(3,6),each=4))
temp=factor(rep(c(20,30,20,30),c(2,2,2,2)))
y=c(2,5,9,11,8,10,16,20)
d=data.frame(time,temp,y)
d
  time temp  y
1    3   20  2
2    3   20  5
3    3   30  9
4    3   30 11
5    6   20  8
6    6   20 10
7    6   30 16
8    6   30 20
```

```
o1=lm(y~time+temp+time:temp,data=d)
o2=lm(y~temp+time+time:temp,data=d)
anova(o1)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
time       1  91.125  91.125 22.0909 0.009308 **
temp       1 120.125 120.125 29.1212 0.005706 **
time:temp  1   3.125   3.125  0.7576 0.433206
Residuals  4  16.500   4.125

anova(o2)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
temp       1 120.125 120.125 29.1212 0.005706 **
time       1  91.125  91.125 22.0909 0.009308 **
temp:time  1   3.125   3.125  0.7576 0.433206
Residuals  4  16.500   4.125
```

```
o3=lm(y~time+temp,data=d)
o4=lm(y~temp+time,data=d)
anova(o3)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
time       1  91.125  91.125  23.217 0.004805 **
temp       1 120.125 120.125  30.605 0.002646 **
Residuals  5  19.625   3.925

anova(o4)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
temp       1 120.125 120.125  30.605 0.002646 **
time       1  91.125  91.125  23.217 0.004805 **
Residuals  5  19.625   3.925
```