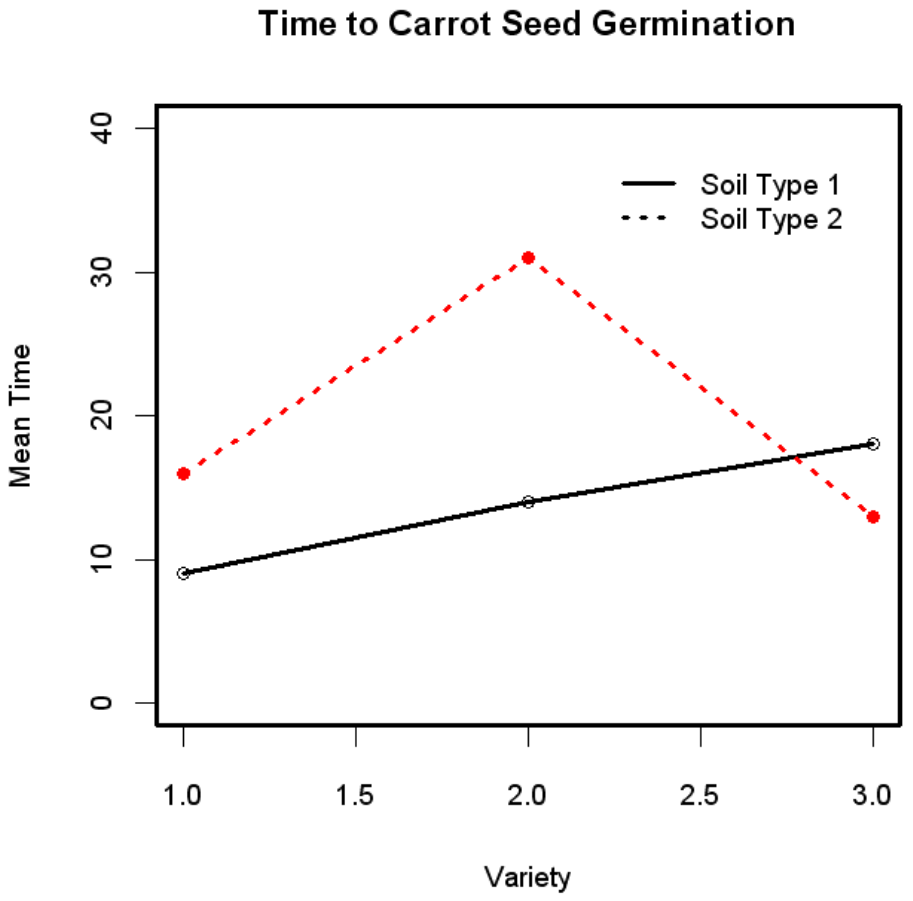
```
> # This file is stored as carrots.r
>
> # First enter the data.  The data file does
> #  not have variable names on the first line
>
> carrot <- read.table("carrots.dat",
+            col.names=c("Soil","Variety","Days"))
> carrot$Soil <- as.factor(carrot$Soil)
> carrot$Variety <- as.factor(carrot$Variety)
> attach(carrot)
>
>
> # Compute sample means of germination times
> # for all combinations of soil type and varieties
> # of carrot seeds.  Make a profile plot
> # At this point UNIX users should open a
> # graphics window with the  motif( ) function
>
> means <- tapply(Days,list(Variety,Soil),mean)
> means
   1  2
1  9 16
2 14 31
3 18 13
>
> #  Set up the axes and title of the profile plot
>
> par(cex=1.2,lwd=3,mex=1.2,mkh=0.15)
> x.axis <- unique(Variety)
> matplot(c(1,3), c(0,40), type="n",
+      xlab="Variety",  ylab="Mean Time",
+      main= "Time to Carrot Seed Germination")
>
> #  Add a profile for each soil type
>
> matlines(x.axis,means,type='l',lty=c(1,3),lwd=3)
>
> # Plot points for the individual observations
>
> matpoints(x.axis,means, pch=c(1,16))
>
> # Add a legend to the plot
>
> legend(2.2,38.6,
+    legend=c('Soil Type 1','Soil Type 2'),
+       lty=c(1,3),bty='n')
```

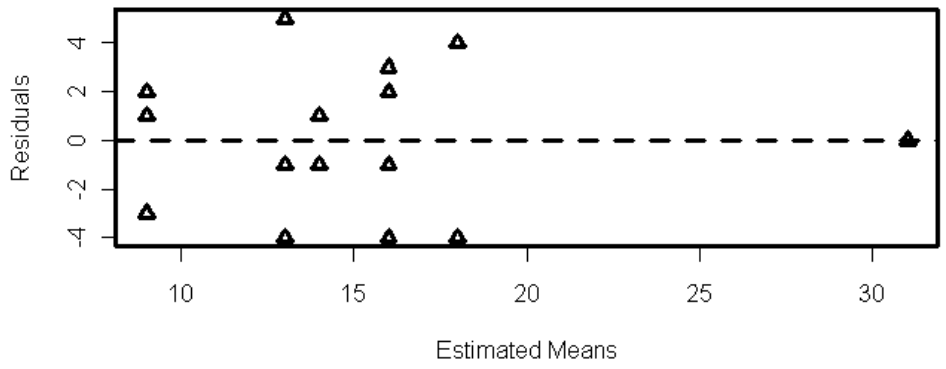


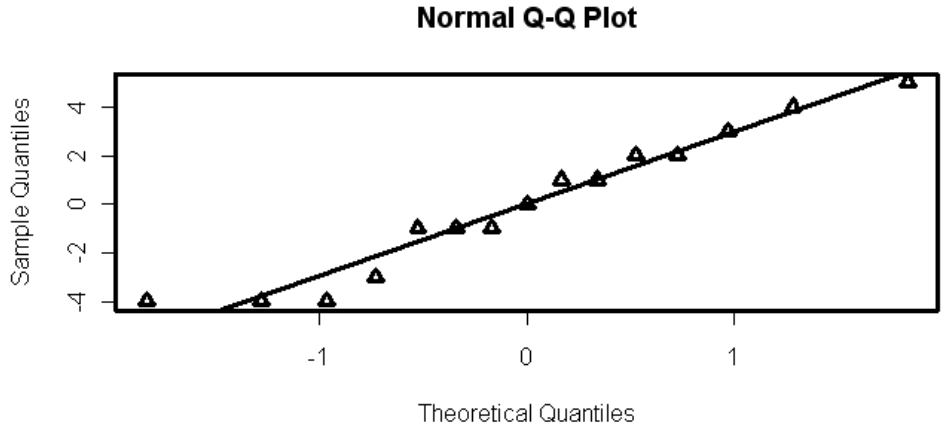
```
> # Fit a model with main effects and interaction
> # Compute both sets of Type I sums of squares
>
> lm.out1 <- lm(Days~Soil*Variety)
> anova(lm.out1)
Analysis of Variance Table

Response: Days
             Df  Sum Sq Mean Sq F value   Pr(>F)
Soil          1  52.500  52.500  3.9375 0.078512 .
Variety       2 124.734  62.367  4.6775 0.040475 *
Soil:Variety  2 222.766 111.383  8.3537 0.008888 **
Residuals     9 120.000  13.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> lm.out2 <- lm(Days~Variety*Soil)
> anova(lm.out2)
Analysis of Variance Table

Response: Days
             Df  Sum Sq Mean Sq F value   Pr(>F)
Variety       2  93.333  46.667  3.5000 0.075085 .
Soil          1  83.901  83.901  6.2926 0.033393 *
Variety:Soil  2 222.766 111.383  8.3537 0.008888 **
Residuals     9 120.000  13.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
>
> # Create a data frame containing the original
> # data and the residuals and estimated means
>
> data.frame(Soil,Variety,Days,Pred=lm.out1$fitted,
+                    Resid=round(lm.out1$resid,3))
   Soil Variety Days Pred Resid
1     1       1    6    9    -3
2     1       1   10    9     1
3     1       1   11    9     2
4     1       2   13   14    -1
5     1       2   15   14     1
6     1       3   14   18    -4
7     1       3   22   18     4
8     2       1   12   16    -4
9     2       1   15   16    -1
10    2       1   19   16     3
11    2       1   18   16     2
12    2       2   31   31     0
13    2       3   18   13     5
14    2       3    9   13    -4
15    2       3   12   13    -1
```

```
> #  Make diagnostic plots.  You must attach the MASS
> #  library to have access to the studres( ) that
> #  computes studentized residuals in the following
> #  code.
>
>  library(MASS)
>
>   par(cex=1.0,mex=1.0,lwd=3,pch=2,mkh=.1)
>   par(mfrow=c(2,1))
>   plot(lm.out1$fitted, lm.out1$resid,
+         xlab="Estimated Means",
+         ylab="Residuals", pch=2, mkh=.1)
>    abline(h=0, lty=2, lwd=3)
>
>    qqnorm(lm.out1$resid)
>    qqline(lm.out1$resid)
```






```
> # By default R uses so-called Helmert contrast
> # matrices for unordered factors and orthogonal
> # polynomial contrast matrices for ordered factors.

>
> options(contrasts=c('contr.helmert','contr.ploy'))
> lm.out <- lm(Days~Soil*Variety)
> model.matrix(lm.out)
```

```
   (Intercept) Soil1 Variety1 Variety2 Soil1:Variety1 Soil1:Variety2
1            1    -1       -1       -1              1              1
2            1    -1       -1       -1              1              1
3            1    -1       -1       -1              1              1
4            1    -1        1       -1             -1              1
5            1    -1        1       -1             -1              1
6            1    -1        0        2              0             -2
7            1    -1        0        2              0             -2
8            1     1       -1       -1             -1             -1
9            1     1       -1       -1             -1             -1
10           1     1       -1       -1             -1             -1
11           1     1       -1       -1             -1             -1
12           1     1        1       -1              1             -1
13           1     1        0        2              0              2
14           1     1        0        2              0              2
15           1     1        0        2              0              2
attr(,"assign")
[1] 0 1 2 2 3 3
attr(,"contrasts")
attr(,"contrasts")$Soil
[1] "contr.helmert"

attr(,"contrasts")$Variety
[1] "contr.helmert"

> summary(lm.out)

Call:
lm(formula = Days ~ Soil * Variety)

Residuals:
       Min         1Q     Median         3Q        Max
-4.000e+00 -2.000e+00 -1.726e-16  2.000e+00  5.000e+00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     16.8333     1.0393  16.196 5.78e-08 ***
Soil1            3.1667     1.0393   3.047  0.01386 *
Variety1         5.0000     1.3176   3.795  0.00425 **
Variety2        -0.6667     0.7082  -0.941  0.37110
Soil1:Variety1   2.5000     1.3176   1.897  0.09027 .
```

```
Soil1:Variety2  -2.8333     0.7082  -4.001  0.00311 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.651 on 9 degrees of freedom
Multiple R-squared: 0.7692,     Adjusted R-squared: 0.641
F-statistic:     6 on 5 and 9 DF,  p-value: 0.01031

> anova(lm.out)
Analysis of Variance Table

Response: Days
             Df  Sum Sq Mean Sq F value   Pr(>F)
Soil          1  52.500  52.500  3.9375 0.078512 .
Variety       2 124.734  62.367  4.6775 0.040475 *
Soil:Variety  2 222.766 111.383  8.3537 0.008888 **
Residuals     9 120.000  13.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



> # The default contrast matrices can be changed by
> # resetting the contrasts options.  The contr.sum
> # option restricts parameters to sum to zero
> # across the levels of any single factor.
>
> options(contrasts=c('contr.sum','contr.ploy'))
>
> # Now, ``contr.sum'' will be used for unordered factors
> # and orthogonal polynomial contrast matrices will
> # be used for ordered factors.
>
> lm.out <- lm(Days~Soil*Variety)
> model.matrix(lm.out)
   (Intercept) Soil1 Variety1 Variety2 Soil1:Variety1 Soil1:Variety2
1            1     1        1        0              1              0
2            1     1        1        0              1              0
3            1     1        1        0              1              0
4            1     1        0        1              0              1
5            1     1        0        1              0              1
6            1     1       -1       -1             -1             -1
7            1     1       -1       -1             -1             -1
8            1    -1        1        0             -1              0
9            1    -1        1        0             -1              0
10           1    -1        1        0             -1              0
11           1    -1        1        0             -1              0
12           1    -1        0        1              0             -1
13           1    -1       -1       -1              1              1
14           1    -1       -1       -1              1              1
15           1    -1       -1       -1              1              1
attr(,"assign")
[1] 0 1 2 2 3 3
attr(,"contrasts")
attr(,"contrasts")$Soil
[1] "contr.sum"

attr(,"contrasts")$Variety
[1] "contr.sum"

> summary(lm.out)

Call:
lm(formula = Days ~ Soil * Variety)

Residuals:
       Min         1Q     Median         3Q        Max
-4.000e+00 -2.000e+00 -1.835e-16  2.000e+00  5.000e+00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    16.8333     1.0393  16.196 5.78e-08 ***
Soil1          -3.1667     1.0393  -3.047  0.01386 *
Variety1       -4.3333     1.3147  -3.296  0.00929 **
Variety2        5.6667     1.6574   3.419  0.00764 **
Soil1:Variety1 -0.3333     1.3147  -0.254  0.80554
```

```
Soil1:Variety2  -5.3333     1.6574  -3.218  0.01052 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.651 on 9 degrees of freedom
Multiple R-squared: 0.7692,     Adjusted R-squared: 0.641
F-statistic:     6 on 5 and 9 DF,  p-value: 0.01031

> anova(lm.out)
Analysis of Variance Table

Response: Days
             Df  Sum Sq Mean Sq F value   Pr(>F)
Soil          1  52.500  52.500  3.9375 0.078512 .
Variety       2 124.734  62.367  4.6775 0.040475 *
Soil:Variety  2 222.766 111.383  8.3537 0.008888 **
Residuals     9 120.000  13.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Compute Type III sums of squares and F-tests.
> # Enter the transpose of the model matrix for
> # the cell means model.
>
>
> # Model matrix for the cell means model
>
> comb <- as.factor(10*as.numeric(Soil) + as.numeric(Variety))
> lm.out <- lm(Days~comb - 1)
> D <- model.matrix(lm.out)
> D
   comb11 comb12 comb13 comb21 comb22 comb23
1       1      0      0      0      0      0
2       1      0      0      0      0      0
3       1      0      0      0      0      0
4       0      1      0      0      0      0
5       0      1      0      0      0      0
6       0      0      1      0      0      0
7       0      0      1      0      0      0
8       0      0      0      1      0      0
9       0      0      0      1      0      0
10      0      0      0      1      0      0
11      0      0      0      1      0      0
12      0      0      0      0      1      0
13      0      0      0      0      0      1
14      0      0      0      0      0      1
15      0      0      0      0      0      1
attr(,"assign")
[1] 1 1 1 1 1 1
attr(,"contrasts")
attr(,"contrasts")$comb
[1] "contr.sum"


>
>
> # Compute the sample means
>
> y <- matrix(Days,ncol=1)
>
> b <- solve(crossprod(D)) %*% crossprod(D,y)
> b
       [,1]
comb11    9
comb12   14
comb13   18
comb21   16
comb22   31
comb23   13
>
> # Generate an identity matrix and a vector of ones
>
> Iden <- function(n) diag(rep(1,n))
> one  <- function(n) matrix(rep(1,n),ncol=1)
>
>
> # Compute Type III sums of squares and
```

```
> # related F-tests
>
> s  <- length(unique(Soil))
> t  <- length(unique(Variety))
>
> yhat <- D %*% b
> sse  <- crossprod(y-yhat)
> df2  <- nrow(y) - s*t
>
>
> # F-test for Type III sum of squares for
> # Soil main effects
>
> c1 <- kronecker( cbind(Iden(s-1),-one(s-1)), t(one(t)) )
> q1 <- t(b) %*% t(c1)%*%
+             solve( c1 %*% solve(crossprod(D)) %*% t(c1))%*%
+             c1 %*% b
> df1<- s-1
> f  <- (q1/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c1
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    1    1   -1   -1   -1
> data.frame(SS=q1,df=df1,F.stat=f,p.value=p)
       SS df   F.stat    p.value
1 123.7714  1 9.282857 0.01386499
>
>
> # F-test for Type III sum of squares for
> # Variety main effects
>
> c2 <- kronecker( t(one(s)), cbind(Iden(t-1),-one(t-1)) )
> q2 <- t(b) %*% t(c2)%*%
+            solve( c2 %*% solve(crossprod(D)) %*% t(c2))%*%
+               c2 %*% b
> df1<- t-1
> f  <- (q2/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c2
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0   -1    1    0   -1
[2,]    0    1   -1    0    1   -1
> data.frame(SS=q2,df=df1,F.stat=f,p.value=p)
       SS df   F.stat    p.value
1 192.1277  2 7.204787 0.01354629
>
>
> # F-test for Type III sum of squares for
> # Soil x Variety interaction effects
>
>
> c3 <- kronecker( cbind(Iden(s-1),-one(s-1)),
+                  cbind(Iden(t-1),-one(t-1)) )
> q3 <- t(b) %*% t(c3)%*%
+              solve( c3 %*% solve(crossprod(D)) %*% t(c3))%*%
+                 c3 %*% b
> df1<- (s-1)*(t-1)
> f  <- (q3/df1)/(sse/df2)
> p  <- 1-pf(f,df1,df2)
> c3
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0   -1   -1    0    1
[2,]    0    1   -1    0   -1    1
> data.frame(SS=q3,df=df1,F.stat=f,p.value=p)
       SS df   F.stat    p.value
1 222.7660  2 8.353723 0.00888845
>
```