STATISTICAL MODELLING

PRACTICAL VIII
SOLUTIONS

VIII.1 A factorial experiment was carried out on a pilot plant scale. A product was being purified by a form of steam distillation process. The five factors, each at 2 levels, were Concentration of material (A), Rate of distillation (B), Volume of solution (C), Stirring rate (D), and Solvent-to-water ratio (E). The residual acidity of material from one run on each of the 32 experimental treatment combinations was determined. The results (in coded form which does not affect the analysis) are given in the following table:

		A_0				A_1			
		D_0		D_1		D_0		D_1	
		E_0	E_1	E_0	E_1	E_0	E_1	E_0	E_1
	C_0	9	3	11	8	10	9	13	7
B_0	C_1	3	5	7	7	5	6	10	7
	C_0	8	4	9	8	6	6	16	6
B_1	C_1	6	4	7	5	10	10	13	6

What are the features of this experiment?

1. Observational unit - a run
2. Response variable - Residual acidity
3. Unrandomized factors - Runs
4. Randomized factors - Conc(A), Rate(B), Vol(C), Stir(D), Solvent(E)
5. Type of study - 2^5 CRD

What is the experimental structure for this experiment?

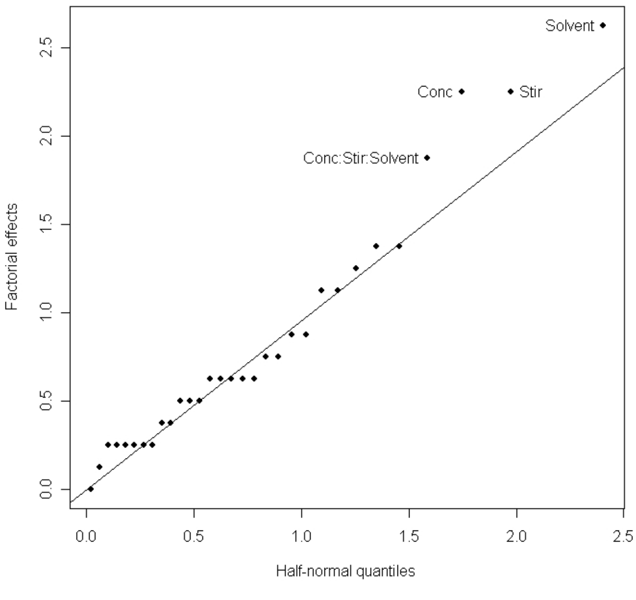
Structure	Formula
unrandomized	*32* Runs
randomized	*2* Conc**2* Rate**2* Vol**2* Stir**2* Solvent

Analyze the data using R, including diagnostic checking. What levels of the factors would you recommend be used to minimize the residual acidity? What residual acidity would be achieved with this (these) combination(s) of the factors?

```
> mp <- c("-", "+")
> fnames <- list(Solvent = mp, Stir = mp, Conc = mp, Vol = mp, Rate = mp)
> Fac5Acid.Treats <- fac.gen(generate = fnames, order="yates")
> Fac5Acid.dat <- data.frame(Runs = factor(1:32), Fac5Acid.Treats)
> remove("Fac5Acid.Treats")
> Fac5Acid.dat$ResAcid <- c(9,3,11,8,10,9,13,7,3,5,7,7,5,6,10,7,
+                           8,4,9,8,6,6,16,6,6,4,7,5,10,10,13,6)
```

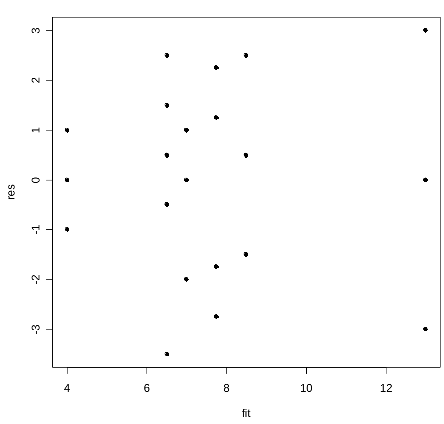
```
> Fac5Acid.dat
   Runs Solvent Stir Conc Vol Rate ResAcid
1      1       -    -    -   -    -       9
2      2       +    -    -   -    -       3
3      3       -    +    -   -    -      11
4      4       +    +    -   -    -       8
5      5       -    -    +   -    -      10
6      6       +    -    +   -    -       9
7      7       -    +    +   -    -      13
8      8       +    +    +   -    -       7
9      9       -    -    -   +    -       3
10    10       +    -    -   +    -       5
11    11       -    +    -   +    -       7
12    12       +    +    -   +    -       7
13    13       -    -    +   +    -       5
14    14       +    -    +   +    -       6
15    15       -    +    +   +    -      10
16    16       +    +    +   +    -       7
17    17       -    -    -   -    +       8
18    18       +    -    -   -    +       4
19    19       -    +    -   -    +       9
20    20       +    +    -   -    +       8
21    21       -    -    +   -    +       6
22    22       +    -    +   -    +       6
23    23       -    +    +   -    +      16
24    24       +    +    +   -    +       6
25    25       -    -    -   +    +       6
26    26       +    -    -   +    +       4
27    27       -    +    -   +    +       7
28    28       +    +    -   +    +       5
29    29       -    -    +   +    +      10
30    30       +    -    +   +    +      10
31    31       -    +    +   +    +      13
32    32       +    +    +   +    +       6
> #
> # analysis
> #
> Fac5Acid.aov <- aov(ResAcid ~ Conc * Rate * Vol * Stir * Solvent
+                        + Error(Runs), Fac5Acid.dat)
> summary(Fac5Acid.aov)
Error: Runs
                           Df   Sum Sq  Mean Sq
Conc                        1   40.500   40.500
Rate                        1    0.500    0.500
Vol                         1   15.125   15.125
Stir                        1   40.500   40.500
Solvent                     1   55.125   55.125
Conc:Rate                   1    2.000    2.000
Conc:Vol                    1    3.125    3.125
Rate:Vol                    1   10.125   10.125
Conc:Stir                   1    0.500    0.500
Rate:Stir                   1    0.500    0.500
Vol:Stir                    1    3.125    3.125
Conc:Solvent                1    3.125    3.125
Rate:Solvent                1    3.125    3.125
Vol:Solvent                 1   12.500   12.500
Stir:Solvent                1   15.125   15.125
Conc:Rate:Vol               1    6.125    6.125
Conc:Rate:Stir              1    2.000    2.000
Conc:Vol:Stir               1    0.125    0.125
Rate:Vol:Stir               1   10.125   10.125
Conc:Rate:Solvent           1    1.125    1.125
Conc:Vol:Solvent            1    0.500    0.500
Rate:Vol:Solvent            1    4.500    4.500
Conc:Stir:Solvent           1   28.125   28.125
Rate:Stir:Solvent           1    1.125    1.125
Vol:Stir:Solvent            1    0.500    0.500
Conc:Rate:Vol:Stir          1    6.125    6.125
```

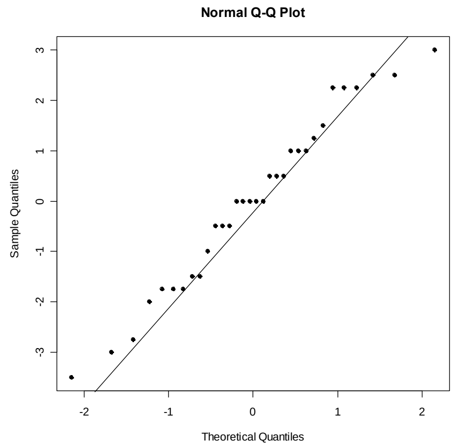
```
Conc:Rate:Vol:Solvent       1     2.000     2.000
Conc:Rate:Stir:Solvent      1     3.125     3.125
Conc:Vol:Stir:Solvent       1     4.500     4.500
Rate:Vol:Stir:Solvent       1     0.500     0.500
Conc:Rate:Vol:Stir:Solvent  1 1.725e-31 1.725e-31
> qqyeffects(Fac5Acid.aov, error.term = "Runs", data=Fac5Acid.dat)
Effect(s) labelled: Conc:Stir:Solvent Conc Stir Solvent
> round(yates.effects(Fac5Acid.aov, error.term="Runs", data=Fac5Acid.dat), 2)
                      Conc                       Rate
                      2.25                       0.25
                       Vol                       Stir
                     -1.37                       2.25
                   Solvent                  Conc:Rate
                     -2.62                       0.50
                  Conc:Vol                   Rate:Vol
                      0.62                       1.12
                 Conc:Stir                  Rate:Stir
                     -0.25                      -0.25
                  Vol:Stir               Conc:Solvent
                     -0.63                      -0.63
              Rate:Solvent                Vol:Solvent
                     -0.62                       1.25
              Stir:Solvent              Conc:Rate:Vol
                     -1.38                       0.88
            Conc:Rate:Stir              Conc:Vol:Stir
                      0.50                      -0.12
             Rate:Vol:Stir          Conc:Rate:Solvent
                     -1.12                      -0.37
          Conc:Vol:Solvent           Rate:Vol:Solvent
                     -0.25                      -0.75
         Conc:Stir:Solvent          Rate:Stir:Solvent
                     -1.87                      -0.37
          Vol:Stir:Solvent         Conc:Rate:Vol:Stir
                     -0.25                      -0.88
     Conc:Rate:Vol:Solvent     Conc:Rate:Stir:Solvent
                      0.50                      -0.63
     Conc:Vol:Stir:Solvent      Rate:Vol:Stir:Solvent
                      0.75                       0.25
Conc:Rate:Vol:Stir:Solvent
                      0.00
```

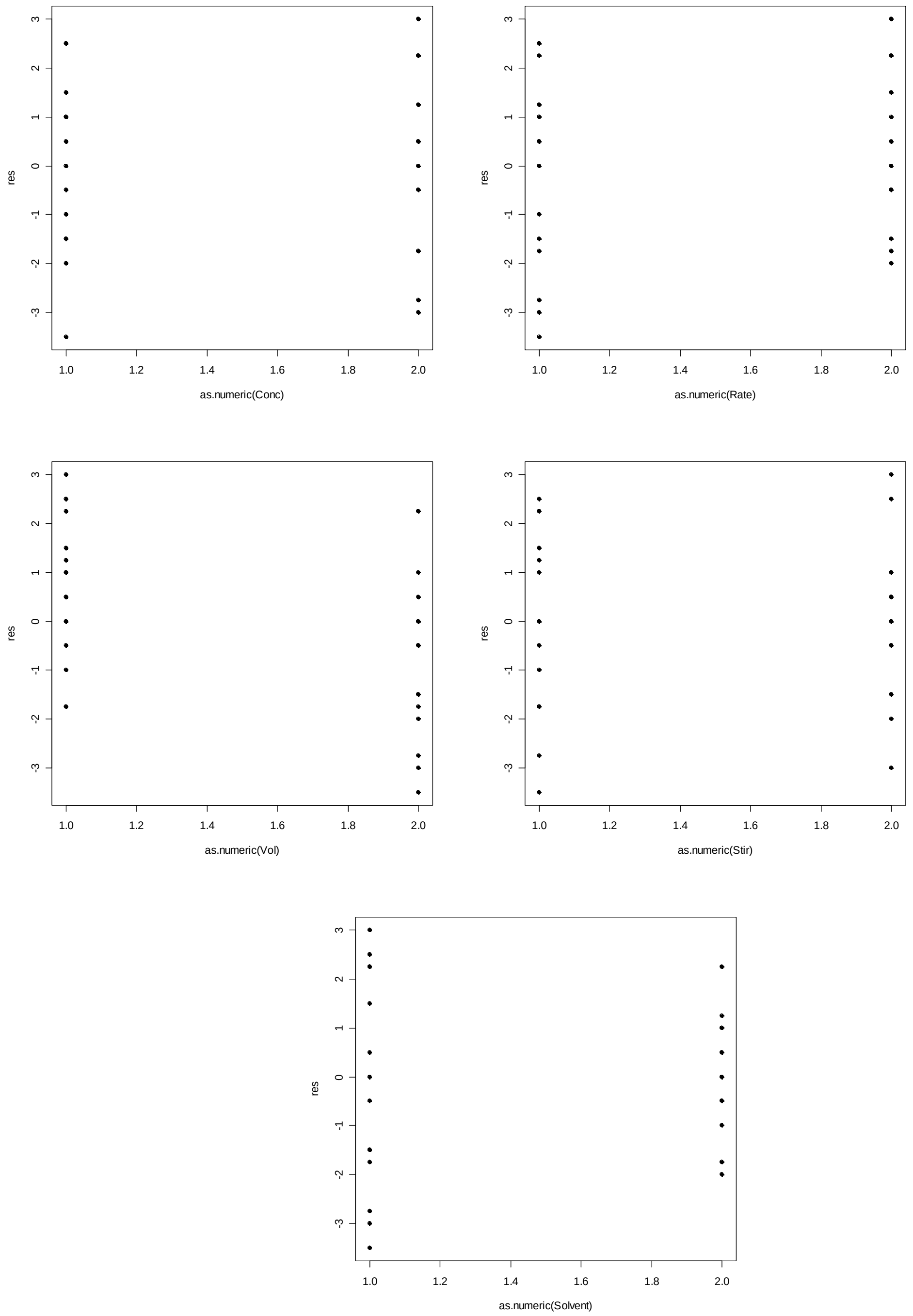


The significant effects appear to be Conc, Stir, Solvent and Conc#Stir#Solvent. We conclude that the three factors Conc, Stir and Solvent interact in their effects on residual acidity. The fitted model is $\psi = E[Y] = \text{Conc}\wedge\text{Stir}\wedge\text{Solvent}$ but the fitted *equation will involve all terms that marginal to this term: any term with one or more of just these factors is marginal to the term. We reanalyse the data for the fitted model and obtain residuals to do the diagnostic checking.*

```
> Fac5Acid.Fit.aov <- aov(ResAcid ~ Conc * Stir * Solvent + Error(Runs),
+                          Fac5Acid.dat)
> summary(Fac5Acid.Fit.aov)

Error: Runs
                  Df Sum Sq Mean Sq F value    Pr(>F)
Conc               1 40.500  40.500 10.5081 0.0034719
Stir               1 40.500  40.500 10.5081 0.0034719
Solvent            1 55.125  55.125 14.3027 0.0009126
Conc:Stir          1  0.500   0.500  0.1297 0.7218629
Conc:Solvent       1  3.125   3.125  0.8108 0.3768278
Stir:Solvent       1 15.125  15.125  3.9243 0.0591628
Conc:Stir:Solvent  1 28.125  28.125  7.2973 0.0124680
Residuals         24 92.500   3.854
> #
> # Diagnostic checking
> #
> tukey.1df(Fac5Acid.Fit.aov, data=Fac5Acid.dat, error.term="Runs")
** Warning - there appears to be extremely little non-linear variation so that
 the values for Tukey.SS are unstable and the results below may be unreliable.
   Only use if at least two non-interacting factors above the same Residual
   in the analysis.
$Tukey.SS
[1] 4.357

$Tukey.F
[1] 1.136914

$Tukey.p
[1] 0.2973707

$Devn.SS
[1] 88.143

> res <- resid.errors(Fac5Acid.Fit.aov)
> fit <- fitted.errors(Fac5Acid.Fit.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> attach(Fac5Acid.dat)
> plot(as.numeric(Conc), res, pch=16)
> plot(as.numeric(Rate), res, pch=16)
> plot(as.numeric(Vol), res, pch=16)
> plot(as.numeric(Stir), res, pch=16)
> plot(as.numeric(Solvent), res, pch=16)
```







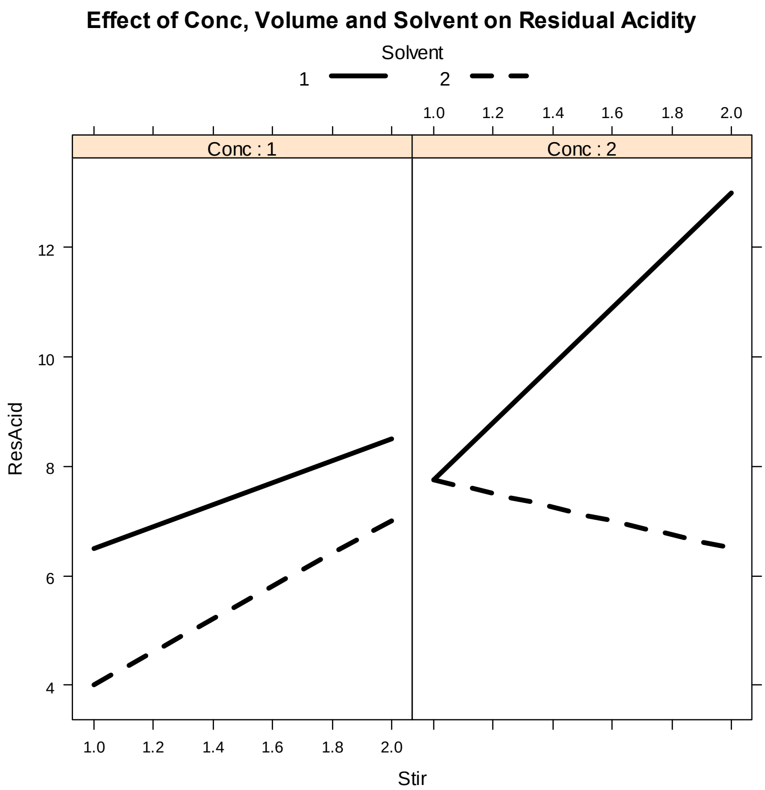
The residual-versus-fitted-values, residuals-versus-factors and normal probability plots all seem satisfactory. Tukey's one-degree-of-freedom-for-nonadditivity is not appropriate for this analysis as it does not involve an additive expectation model.

The table of means for the three-factor interaction is given in the following output and the corresponding interaction plot is also included. Tukey's HSD is computed to enable one to decide which means are significantly different. Examination of the table of means reveals that the combination of the low levels of Conc and Stir with the high level of Solvent (- - +) produces the lowest residual acidity. However, the only significant differences are between this combination and the combination that has Concentration and Stir at their high *levels and the low level of Solvent (+ + -). Consequently, any combination except this last one (+ + -) could be used as there is no evidence of any difference between those combinations.*

```
> #
> # treatment differences
> #
> interaction.ABC.plot(ResAcid, Stir, Solvent, Conc, data=Fac5Acid.dat,
+             title="Effect of Conc, Volume and Solvent on Residual Acidity")
> Fac5Acid.means <- model.tables(Fac5Acid.Fit.aov, type="means")
> Fac5Acid.means$tables$"Conc:Stir:Solvent"
, , Solvent = -

    Stir
Conc -     +
   -  6.50  8.50
   +  7.75 13.00

, , Solvent = +

    Stir
Conc -     +
   -  4.00  7.00
   +  7.75  6.50

> q <- qtukey(0.95, 8, 24)
> q
[1] 4.683752
```

So Tukey's HSD is

$$w\left(5\%\right)=\frac{4.683752}{\sqrt{2}}\times\sqrt{\frac{3.854\times 2}{4}}=4.60$$



VIII.2 A new rifle was being tested for performance to decide some characteristics of the weapon. The testing programme involved a four-factor factorial experiment consisting of 8 tests run over two days as only 8 tests could be run on a single day. It was decided to confound the four-factor interaction with the day difference.

The four factors to be investigated were the propellant charge, the weight of the projectile, the propellant web and two different weapons of the type being evaluated. The velocity of the projectiles was measured and the results were as follows:

Day	Test	Charge Weight	Projectile Weight	Propellant Web	Weapon	Velocity
	1	1	1	1	1	197
	2	2	2	1	1	250
	3	1	2	2	1	115
	4	2	1	2	1	200
1	5	1	2	1	2	153
	6	2	1	1	2	245
	7	1	1	2	2	126
	8	2	2	2	2	154
	1	1	2	1	1	168
	2	2	1	1	1	251
	3	1	1	2	1	139
	4	2	2	2	1	166
2	5	1	1	1	2	175
	6	2	2	1	2	241
	7	1	2	2	2	84
	8	2	1	2	2	197

What are the features of this experiment?

1. Observational unit - a test
2. Response variable - Velocity
3. Unrandomized factors - Days, Tests
4. Randomized factors - Charge, Project, Propell, Weapon
5. Type of study - confounded 2^4 RCBD

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	*2* Days/*8* Tests
randomized	*2* Charge**2* Project**2* Propell**2* Weapon

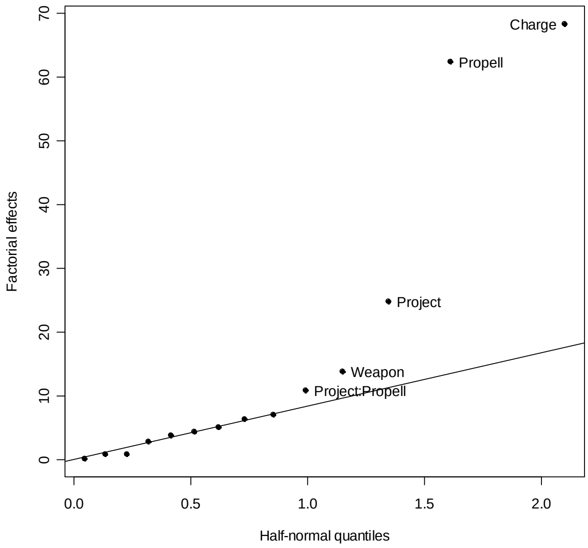
What are the expected mean squares for the lines in the analysis of variance table based on all unrandomized factors being random and all randomized factors being fixed?

The expected mean squares from the unrandomized structure are just those for the randomized complete block design — that for contribution for Day is $\sigma^2_{\mathrm{DT}} + 8\sigma^2_{\mathrm{D}}$ *and for Tests[Day] is* σ^2_{DT}*. The contributions from the randomized structure are just the q functions with appropriate subscripts.*

Source	df	E[MSq]
Day	1	
Charge#Project#Propell#Weapon	1	$\sigma^2_{\mathrm{DT}} + 8\sigma^2_{\mathrm{D}} + q_{\mathrm{CJPW}}(\boldsymbol{\psi})$
Tests[Day]		
Charge	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{C}}(\boldsymbol{\psi})$
Project	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{J}}(\boldsymbol{\psi})$
Propell	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{P}}(\boldsymbol{\psi})$
Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{W}}(\boldsymbol{\psi})$
Charge#Project	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CJ}}(\boldsymbol{\psi})$
Charge#Propell	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CP}}(\boldsymbol{\psi})$
Project#Propell	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{JP}}(\boldsymbol{\psi})$
Charge#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CW}}(\boldsymbol{\psi})$
Project#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{JW}}(\boldsymbol{\psi})$
Propell#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{PW}}(\boldsymbol{\psi})$
Charge#Project#Propell	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CJP}}(\boldsymbol{\psi})$
Charge#Project#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CJW}}(\boldsymbol{\psi})$
Charge#Propell#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{CPW}}(\boldsymbol{\psi})$
Project#Propell#Weapon	1	$\sigma^2_{\mathrm{DT}} + q_{\mathrm{JPW}}(\boldsymbol{\psi})$

Analyze the data using R, including diagnostic checking. What levels of the factors would you recommend be used to maximize the velocity? What velocity would be achieved with this (these) combination(s) of the factors?

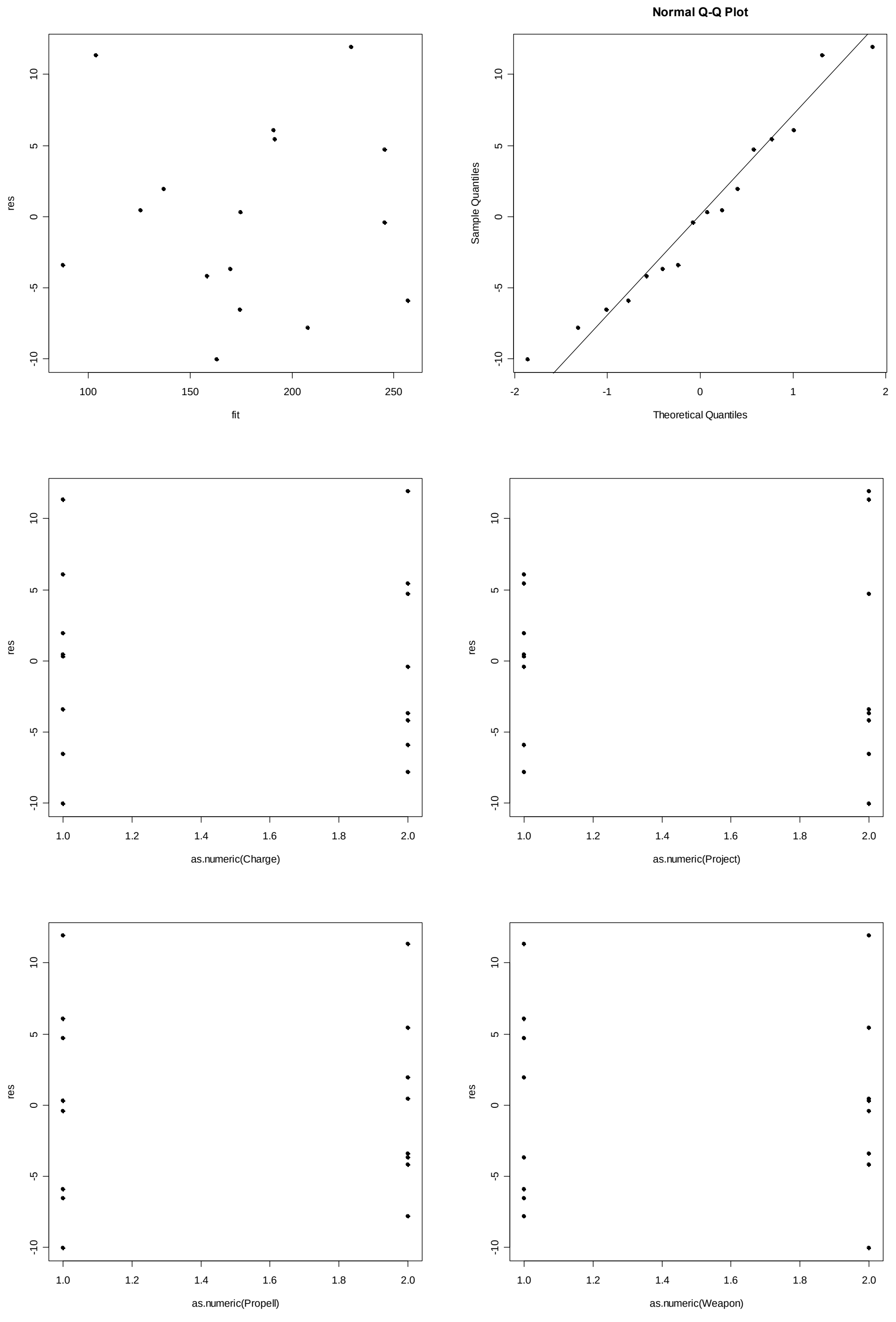
```
> Fac4Ball.dat
   Day Tests Charge Project Propell Weapon Velocity
1    1     1      -       -       -      -      197
2    1     2      +       +       -      -      250
3    1     3      -       +       +      -      115
4    1     4      +       -       +      -      200
5    1     5      -       +       -      +      153
6    1     6      +       -       -      +      245
7    1     7      -       -       +      +      126
8    1     8      +       +       +      +      154
9    2     1      -       +       -      -      168
10   2     2      +       -       -      -      251
11   2     3      -       -       +      -      139
12   2     4      +       +       +      -      166
13   2     5      -       -       -      +      175
14   2     6      +       +       -      +      241
15   2     7      -       +       +      +       84
16   2     8      +       -       +      +      197
> #
> # analysis
> #
> Fac4Ball.aov <- aov(Velocity ~ Charge * Project * Propell * Weapon +
+                            Error(Day/Tests), Fac4Ball.dat)
> summary(Fac4Ball.aov)

Error: Day
                              Df Sum Sq Mean Sq
Charge:Project:Propell:Weapon  1 22.562  22.562

Error: Day:Tests
                       Df  Sum Sq Mean Sq
Charge                  1 18700.6 18700.6
Project                 1  2475.1  2475.1
Propell                 1 15562.6 15562.6
Weapon                  1   770.1   770.1
Charge:Project          1    76.6    76.6
Charge:Propell          1   105.1   105.1
Project:Propell         1   473.1   473.1
Charge:Weapon           1   162.6   162.6
Project:Weapon          1    33.1    33.1
Propell:Weapon          1     3.1     3.1
Charge:Project:Propell  1   203.1   203.1
Charge:Project:Weapon   1     0.1     0.1
Charge:Propell:Weapon   1     3.1     3.1
Project:Propell:Weapon  1    60.1    60.1
> qqyeffects(Fac4Ball.aov, error.term = "Day:Tests", data=Fac4Ball.dat)
Effect(s) labelled: Project:Propell Weapon Project Propell Charge
> round(yates.effects(Fac4Ball.aov, error.term="Day:Tests",
+                                                   data=Fac4Ball.dat), 2)
                Charge                Project                Propell
                 68.38                 -24.87                 -62.37
                Weapon         Charge:Project         Charge:Propell
                -13.87                   4.37                  -5.12
       Project:Propell          Charge:Weapon         Project:Weapon
                -10.88                   6.37                  -2.88
        Propell:Weapon Charge:Project:Propell  Charge:Project:Weapon
                 -0.88                  -7.12                  -0.12
 Charge:Propell:Weapon Project:Propell:Weapon
                  0.88                  -3.87
```



The normal plot indicates that the four main effects are significant and that Project#Propell interaction may need to be taken into account. Thus the fitted model would appear to be $\psi = E[Y] = \text{Charge} + \text{Weapon} + \text{Project} \wedge \text{Propell}$.

```
> Fac4Ball.Fit.aov <- aov(Velocity ~ Day + Charge + Weapon + Project * Propell
+
+                                 Error(Day/Tests), Fac4Ball.dat)
> summary(Fac4Ball.Fit.aov)

Error: Day
    Df Sum Sq Mean Sq
Day  1 22.562  22.562

Error: Day:Tests
                Df  Sum Sq Mean Sq  F value    Pr(>F)
Charge           1 18700.6 18700.6 260.3075 5.981e-08
Weapon           1   770.1   770.1  10.7191 0.0096210
Project          1  2475.1  2475.1  34.4523 0.0002379
Propell          1 15562.6 15562.6 216.6273 1.330e-07
Project:Propell  1   473.1   473.1   6.5849 0.0303811
Residuals        9   646.6    71.8
> #
> # Diagnostic checking
> #
> tukey.1df(Fac4Ball.Fit.aov, data=Fac4Ball.dat, error.term="Day:Tests")
$Tukey.SS
[1] 19.52812

$Tukey.F
[1] 0.2491489

$Tukey.p
[1] 0.6311098

$Devn.SS
[1] 627.0344

> res <- resid.errors(Fac4Ball.Fit.aov)
> fit <- fitted.errors(Fac4Ball.Fit.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> attach(Fac4Ball.dat)
> plot(as.numeric(Charge), res, pch=16)
```

```
> plot(as.numeric(Project), res, pch=16)
> plot(as.numeric(Propell), res, pch=16)
> plot(as.numeric(Weapon), res, pch=16)
```

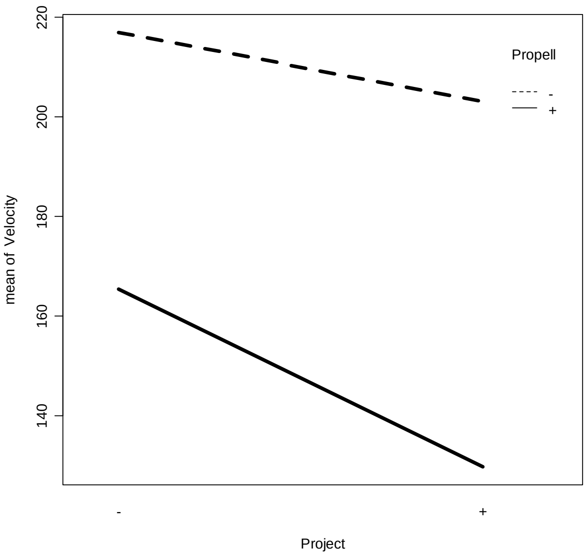


The residuals-versus-fitted values, residuals-versus-factor and normal probability plots for the fitted model are all satisfactory. Tukey's one-degree-of – freedom-for-nonadditivity is not significant.

```
> #
> # treatment differences
> #
> interaction.plot(Project, Propell, Velocity, lwd=4)
> Fac4Ball.means <- model.tables(Fac4Ball.Fit.aov, type="means")
> Fac4Ball.means$tables$"Grand mean"
[1] 178.8125
> Fac4Ball.means$tables$"Charge"
Charge
      -       +
144.625 213.000
> Fac4Ball.means$tables$"Weapon"
Weapon
      -       +
185.750 171.875
> Fac4Ball.means$tables$"Project:Propell"
       Propell
Project -      +
      - 217.00 165.50
      + 203.00 129.75
> q <- qtukey(0.95, 4, 9)
> q
[1] 4.41489
```

So Tukey's HSD is

$$w(5\%) = \frac{4.41489}{\sqrt{2}} \times \sqrt{\frac{71.8 \times 2}{4}} = 18.70$$



The combination that maximizes the velocity is Charge at the high level, Weapon at the low level, Propell at the low level and Project at either level. The latter is the case because there was not a significant difference between the Project means with Propell at the low level. On the other hand there is a significant difference between Propell means at the high level of Project and almost a significant difference at the low level of Project — Propell at the low level gives the higher velocity.

The velocity that would be achieved with these combinations can be computed using the following equation for the response:

$$E[Y] = 178.8125 + \frac{68.375}{2} x_{\text{Charge}} - \frac{13.875}{2} x_{\text{Weapon}} - \frac{24.875}{2} x_{\text{Proj}} - \frac{62.375}{2} x_{\text{Prop}} - \frac{10.875}{2} x_{\text{Proj}} x_{\text{Prop}}$$

where x_{Charge}, x_{Weapon}, x_{Proj} and x_{Prop} take the values ± 1 according as to whether the low or high level of the corresponding factor is involved.

The fitted values for the two recommended combinations are:

$$\begin{aligned} E[Y] &= 178.8125 + \frac{68.375}{2}(1) - \frac{13.875}{2}(-1) \\ &\quad - \frac{24.875}{2}(-1) - \frac{62.375}{2}(-1) - \frac{10.875}{2}(-1)(-1) \\ &= 178.8125 + 34.1875 + 6.9375 + 12.4375 + 31.1875 - 5.4375 \\ &= 258.125 \end{aligned}$$

$$\begin{aligned} y &= 178.8125 + \frac{68.375}{2}(1) - \frac{13.875}{2}(-1) \\ &\quad - \frac{24.875}{2}(1) - \frac{62.375}{2}(-1) - \frac{10.875}{2}(1)(-1) \\ &= 178.8125 + 34.1875 + 6.9375 - 12.4375 + 31.1875 + 5.4375 \\ &= 244.125 \end{aligned}$$

VIII.3 A processing experiment is to be run to investigate the effects of 6 factors, each at two levels, on the total yield of peanut oil from batches of peanuts. To save on resources the experimenter decides to use a quarter of the complete set of treatment combinations. Use the table given in subsection e) of section X.D, *Fractional factorial design at two levels*, to identify a suitable design.

a) What is the resolution of this design?

The design is a 2^{6-2}_{IV} *and so is of resolution IV.*

b) What are the implications of the design's resolution?

Being of resolution IV means that main effects are aliased with three factor interactions and tow-factor interactions are aliased with two-factor interactions.

c) What are the generators and defining relations for the design?

From the table the generators for the design are I = ABCE = BCDF.

Consequently the defining relations are: I = ABCE = BCDF = ADEF.

d) What is its aliasing pattern?

The aliasing pattern is obtained by multiplying all effects by the defining relations. It is given in the following table.

I + ABCE + ADEF + BCDF
A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF

e) What treatment combinations should the experimenter include in the experiment?

A	B	C	D	E	F
-	-	-	-	-	-
+	-	-	-	+	-
-	+	-	-	+	+
+	+	-	-	-	+
-	-	+	-	+	+
+	-	+	-	-	+
-	+	+	-	-	-
+	+	+	-	+	-
-	-	-	+	-	+
+	-	-	+	+	+
-	+	-	+	+	-
+	+	-	+	-	-
-	-	+	+	+	-
+	-	+	+	-	-
-	+	+	+	-	+
+	+	+	+	+	+

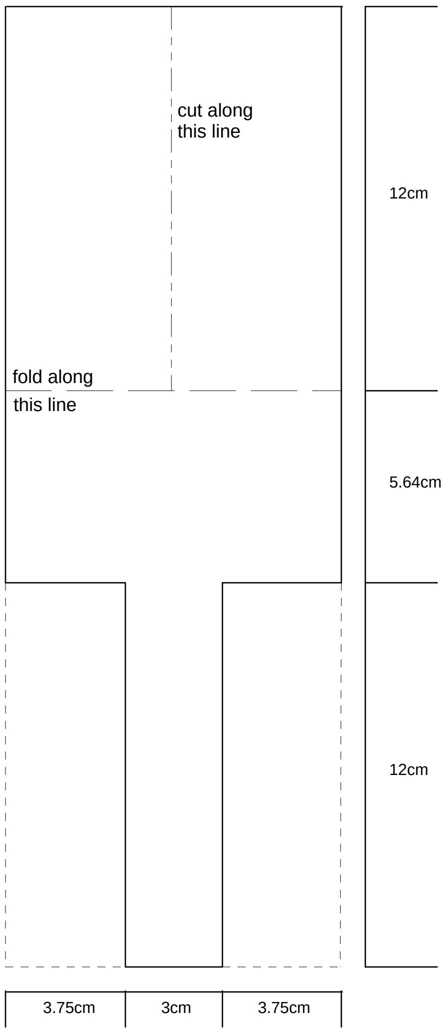
VIII.4 An experimenter wants to investigate 5 factors at 2 levels but has only enough resources for 8 runs. Use R to obtain a randomized layout for the experimenter using a seed of 124.

To get 8 runs with 5 factors requires a 2^{5-2} design and from the table in subsection e) of VIII.D, we find that the generators for a 2_{III}^{5-2} design are D = AB and E = AC. The R expressions for the layout, and their output, are as follows:

```
> #
> # set up randomized factors
> #
> mp <- c("-", "+")
> Frf5.2.ran <- fac.gen(generate = list(A = mp, B = mp, C = mp), order="yates")
> attach(Frf5.2.ran)
> Frf5.2.ran$D <- factor(mpone(A)*mpone(B), labels = mp)
> Frf5.2.ran$E <- factor(mpone(A)*mpone(C), labels = mp)
> detach(Frf5.2.ran)
> #
> # randomize
> #
> n <- 8
> Frf5.2.unit <- list(Runs = n)
> Frf5.2.lay <- fac.layout(unrandomized = Frf5.2.unit, randomized = Frf5.2.ran,
+                          seed = 124)
> Frf5.2.lay
  Units Permutation Runs A B C D E
1     1           1    1 - - - + +
2     2           5    2 - - + + -
3     3           7    3 + - + - +
4     4           4    4 + + - + -
5     5           2    5 + - - - -
6     6           3    6 + + + + +
7     7           8    7 - + - - +
8     8           6    8 - + + - -
```

VIII.5 The Light Helicopter Corporation wishes to investigate ways in which the flight time of their helicopters can be increased. The standard design for the helicopters they produce in shown below.

The standard design



Improving the design

Engineers from their company have got together and had a brainstorming session to identify modifications to the design that might increase the flight time. They suggested that the following factors be investigated.

Factors		-	+
Paper type	(P)	light	heavy
Wing length	(W)	7.5cm	12cm
Body length	(L)	7.5cm	12cm
Body width	(B)	3cm	5cm
Paper clip	(C)	no	yes
Fold	(F)	no	yes
Taped body	(T)	no	yes
Taped wing	(M)	no	yes

Now there are 8 factors to be investigated. If all combinations of the factors were to be investigated, as in a complete factorial, how many helicopters would have to be produced?

It is decided that the full set cannot be run and that a fractional factorial must be employed. There are sufficient resources to make 16 helicopters at this stage. To study the 8 factors in 16 runs a 2^{8-4}_{IV} fractional factorial design is chosen. The design has generators **5** = **234**, **6** = **134**, **7** = **123** and **8** = **124**. The runs, given in standard order, are given in the following table:

	Factor							
Standard Order	**1 P**	**2 W**	**3 L**	**4 B**	**5 C**	**6 F**	**7 T**	**8 M**
1	-	-	-	-	-	-	-	-
2	+	-	-	-	-	+	+	+
3	-	+	-	-	+	-	+	+
4	+	+	-	-	+	+	-	-
5	-	-	+	-	+	+	+	-
6	+	-	+	-	+	-	-	+
7	-	+	+	-	-	+	-	+
8	+	+	+	-	-	-	+	-
9	-	-	-	+	+	+	-	+
10	+	-	-	+	+	-	+	-
11	-	+	-	+	-	+	+	-
12	+	+	-	+	-	-	-	+
13	-	-	+	+	-	-	+	+
14	+	-	+	+	-	+	-	-
15	-	+	+	+	+	-	-	-
16	+	+	+	+	+	+	+	+

The aliasing pattern (ignoring three- and more-factor interactions and substituting in factor names) for this experiment is as follows:

ℓ_1 ➔ average	ℓ_0 ➔ average
ℓ_2 ➔ **1**	ℓ_P ➔ **P**
ℓ_3 ➔ **2**	ℓ_W ➔ **W**
ℓ_4 ➔ **12** + **37** + **48** + **56**	ℓ_{PW} ➔ **PW** + **LT** + **BM** + **CF**
ℓ_5 ➔ **3**	ℓ_L ➔ **L**
ℓ_6 ➔ **13** + **27** + **46** + **58**	ℓ_{PL} ➔ **PL** + **WT** + **BF** + **CM**
ℓ_7 ➔ **23** + **17** + **45** + **68**	ℓ_{WL} ➔ **WL** + **PT** + **BC** + **FM**
ℓ_8 ➔ **7**	ℓ_T ➔ **T**
ℓ_9 ➔ **4**	ℓ_B ➔ **B**
ℓ_{10} ➔ **14** + **28** + **36** + **57**	ℓ_{PB} ➔ **PB** + **WM** + **LF** + **CT**
ℓ_{11} ➔ **24** + **18** + **35** + **67**	ℓ_{WB} ➔ **WB** + **PM** + **LC** + **FT**
ℓ_{12} ➔ **8**	ℓ_M ➔ **M**
ℓ_{13} ➔ **34** + **16** + **25** + **78**	ℓ_{LB} ➔ **LB** + **PF** + **WC** + **TM**
ℓ_{14} ➔ **6**	ℓ_F ➔ **F**
ℓ_{15} ➔ **5**	ℓ_C ➔ **C**
ℓ_{16} ➔ **15** + **26** + **38** + **47**	ℓ_{PF} ➔ **PC** + **WF** + **LM** + **BT**

Generators:

C = **WLB**, **F** = **PLB**, **T** = **PWL** and **M** = **PWB**.

Analysis of results

What is the experimental structure for this experiment?

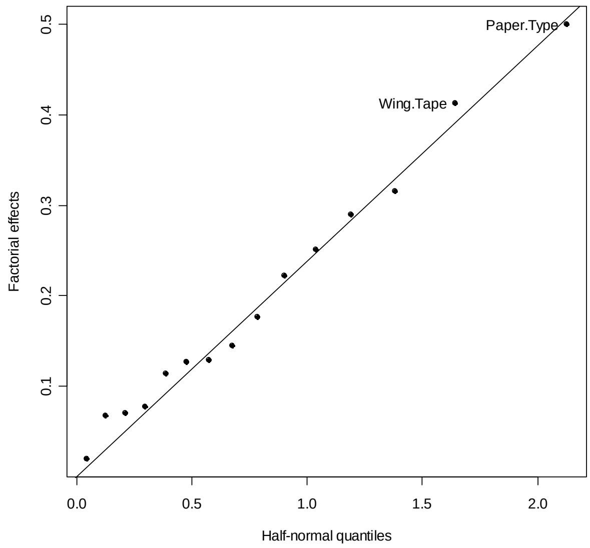
Structure	Formula
unrandomized	Runs
randomized	*2* P**2* W**2* L**2* B**2* C**2* F**2* T**2* M

Use R to analyse the results of the experiment and to perform appropriate diagnostic checking. What treatment combinations would give the longest flight time and what would you predict would be the flight time for these treatment combinations? The treatment combinations are available from the *Computing files* page of the web site in the file *Frf8Heli.Desgn.sdd*.

The following R output contains the analysis of the times recorded.

```
> Frf8Heli.2003.dat
   Standard.Order Runs Paper.Type Wing.Length Body.Length Body.Width Clip
 4              4    1          +           +           -          -    +
 3              3    2          -           +           -          -    +
14             14    3          +           -           +          +    -
 6              6    4          +           -           +          -    +
 8              8    5          +           +           +          -    -
13             13    6          -           -           +          +    -
 7              7    7          -           +           +          -    -
 2              2    8          +           -           -          -    -
11             11    9          -           +           -          +    -
16             16   10          +           +           +          +    +
 1              1   11          -           -           -          -    -
 5              5   12          -           -           +          -    +
15             15   13          -           +           +          +    +
12             12   14          +           +           -          +    -
 9              9   15          -           -           -          +    +
10             10   16          +           -           -          +    +

   Fold Body.Tape Wing.Tape Time.Giang Time.Alan  Time
 4    +         -         -       2.17      1.86 2.015
 3    -         +         +       1.43      1.22 1.325
14    +         -         -       1.79      1.36 1.575
 6    -         -         +       1.63      1.41 1.520
 8    -         +         -       2.14      1.79 1.965
13    -         +         +       2.41      2.22 2.315
 7    +         -         +       1.86      2.02 1.940
 2    +         +         +       1.44      0.99 1.215
11    +         +         -       2.64      2.49 2.565
16    +         +         +       1.88      1.78 1.830
 1    -         -         -       3.07      2.76 2.915
 5    +         +         -       2.03      2.01 2.020
15    -         -         -       2.76      2.32 2.540
12    -         -         +       1.93      1.71 1.820
 9    +         -         +       1.83      1.78 1.805
10    -         +         -       1.68      1.28 1.480
> #
> # analyse
> #
> Frf8Heli.2003.aov <- aov(Time ~ (Paper.Type + Wing.Length + Body.Length +
+          Body.Width + Clip + Fold + Body.Tape + Wing.Tape)^2 + Error(Runs),
+                                                       Frf8Heli.2003.dat)
> summary(Frf8Heli.2003.aov)

Error: Runs
                       Df  Sum Sq Mean Sq
Paper.Type              1 1.00250 1.00250
Wing.Length             1 0.08338 0.08338
Body.Length             1 0.01995 0.01995
Body.Width              1 0.06439 0.06439
Clip                    1 0.19691 0.19691
Fold                    1 0.05233 0.05233
Body.Tape               1 0.12514 0.12514
Wing.Tape               1 0.68269 0.68269
Paper.Type:Wing.Length  1 0.39848 0.39848
Paper.Type:Body.Length  1 0.00150 0.00150
Paper.Type:Body.Width   1 0.06695 0.06695
Paper.Type:Clip         1 0.33495 0.33495
Paper.Type:Fold         1 0.02364 0.02364
Paper.Type:Body.Tape    1 0.01789 0.01789
Paper.Type:Wing.Tape    1 0.25125 0.25125
> qqyeffects(Frf8Heli.2003.aov, error.term = "Runs", data=Frf8Heli.2003.dat)
Effect(s) labelled: Wing.Tape Paper.Type
> round(yates.effects(Frf8Heli.2003.aov, error.term="Runs",
data=Frf8Heli.2003.dat), 2)
            Paper.Type           Wing.Length            Body.Length
                 -0.50                  0.14                   0.07
            Body.Width                  Clip                   Fold
```

```
                 0.13                       -0.22                      -0.11
            Body.Tape                   Wing.Tape  Paper.Type:Wing.Length
                -0.18                       -0.41                       0.32
Paper.Type:Body.Length  Paper.Type:Body.Width          Paper.Type:Clip
                 0.02                       -0.13                       0.29
       Paper.Type:Fold   Paper.Type:Body.Tape      Paper.Type:Wing.Tape
                 0.08                        0.07                       0.25
```



*It would appear that there are two significant main effects: Paper.Type (**P**) and Wing.Tape (**M**). The most likely fitted model is*

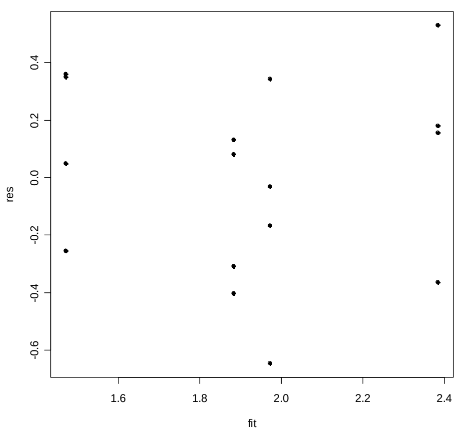
$$\psi = E[Y] = Paper.Type + Wing.Tape$$

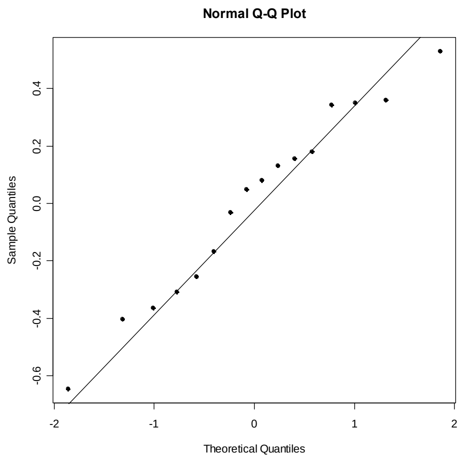
The significant terms have been fitted and diagnostic checking done on the residuals produced.

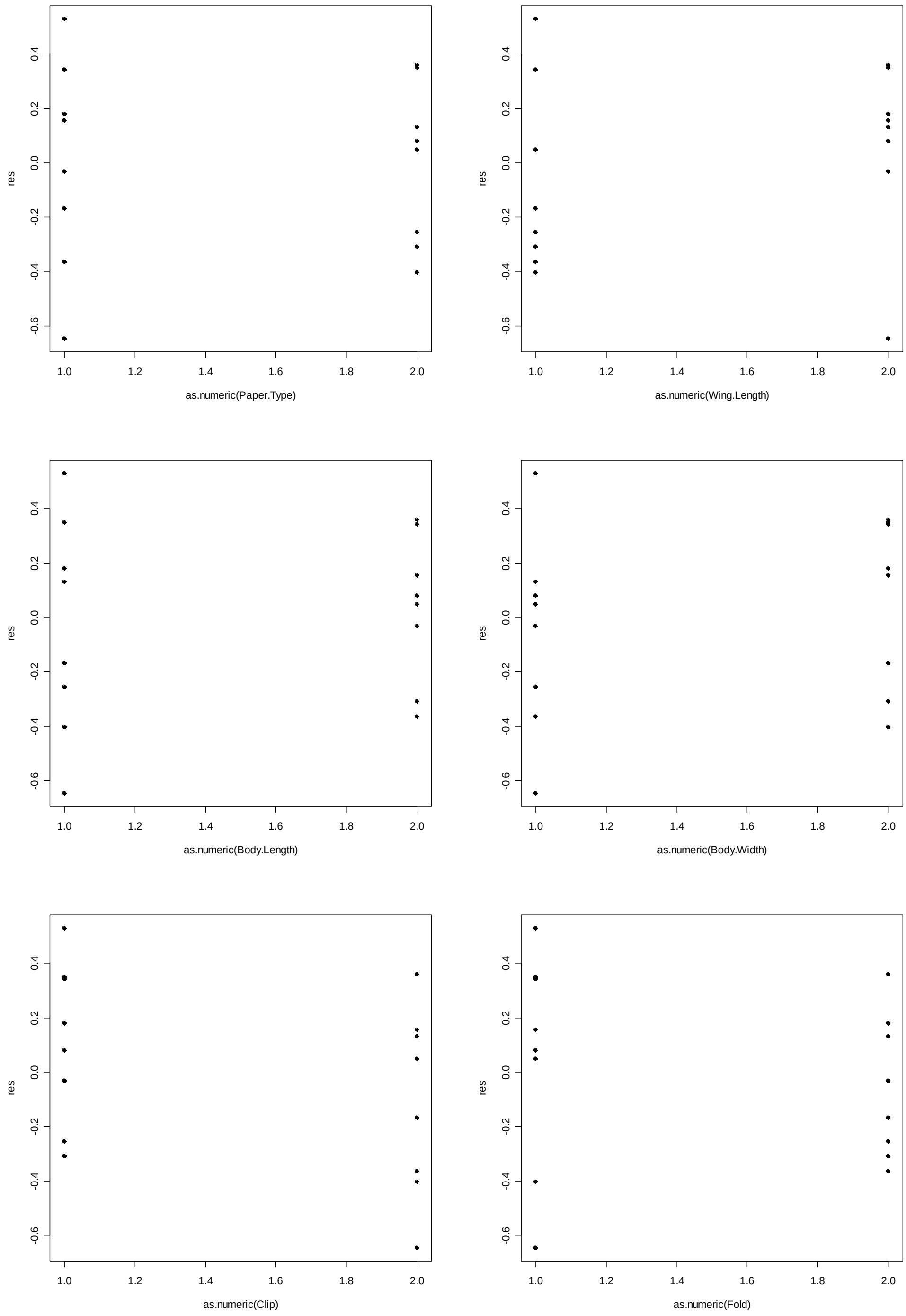
```
> Frf8Heli.2003.Fit.aov <- aov(Time ~ Paper.Type+Wing.Tape + Error(Runs),
+                                                        Frf8Heli.2003.dat)
> summary(Frf8Heli.2003.Fit.aov)

Error: Runs
           Df  Sum Sq Mean Sq F value  Pr(>F)
Paper.Type  1 1.00250 1.00250  7.9624 0.01442
Wing.Tape   1 0.68269 0.68269  5.4223 0.03666
Residuals  13 1.63676 0.12590
```

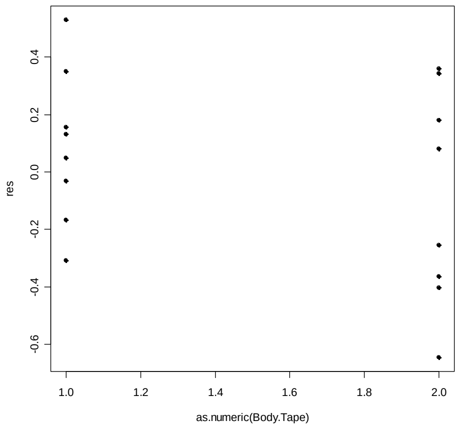
It would appear that the variability of the results was relatively low. An estimate of the variability is provided by the $\sqrt{\text{Residual MSq}}$ *from the analysis. That is,* $s = \sqrt{0.1259} = 0.35$*. So one can expect repeat runs with the same configuration to differ by as much as 0.35 of a second. This compares favourably with the previous values between 0.31 and 0.40.*

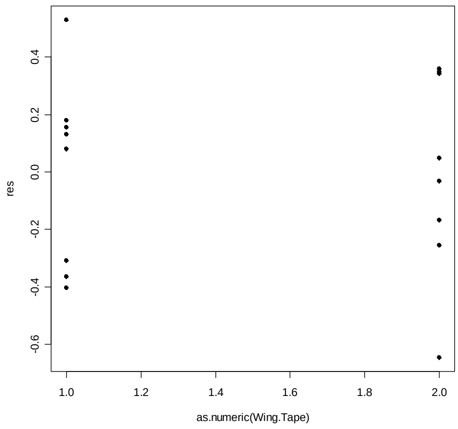
```
>
> #
> # Diagnostic checking
> #
> tukey.1df(Frf8Heli.2003.Fit.aov, data = Frf8Heli.2003.dat,
error.term="Runs")
$Tukey.SS
[1] 0.2512516

$Tukey.F
[1] 2.176113

$Tukey.p
[1] 0.1659172

$Devn.SS
[1] 1.385506

> res <- resid.errors(Frf8Heli.2003.Fit.aov)
> fit <- fitted.errors(Frf8Heli.2003.Fit.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> attach(Frf8Heli.2003.dat)
> plot(as.numeric(Paper.Type), res, pch=16)
> plot(as.numeric(Wing.Length), res, pch=16)
> plot(as.numeric(Body.Length), res, pch=16)
> plot(as.numeric(Body.Width), res, pch=16)
> plot(as.numeric(Clip), res, pch=16)
> plot(as.numeric(Fold), res, pch=16)
> plot(as.numeric(Body.Tape), res, pch=16)
> plot(as.numeric(Wing.Tape), res, pch=16)
```






res
as.numeric(Paper.Type)
as.numeric(Wing.Length)
as.numeric(Body.Length)
as.numeric(Body.Width)
as.numeric(Clip)
as.numeric(Fold)
-0.6
-0.4
-0.2
0.0
0.2
0.4
1.0
1.2
1.4
1.6
1.8
2.0





The residuals-versus-fitted-values and residuals-versus-factors plots appear to be satisfactory except for a single outlier. So the homogeneity of variance assumption seems to be met. Tukey's one-degree-of-freedom is not significant so that there is no evidence of nonadditivity. The normal probability plot displays a roughly straight-line pattern and so the normality assumption appears to be met.

The tables of means to be used in summarizing the results of the experiment are as follows:

```
> #
> # treatment differences
> #
> Frf8Heli.2003.means <- model.tables(Frf8Heli.2003.Fit.aov, type="means")
> Frf8Heli.2003.means$tables$"Grand mean"
[1] 1.927813
> Frf8Heli.2003.means$tables$"Paper.Type"
Paper.Type
       -        +
2.178125 1.677500
> Frf8Heli.2003.means$tables$"Wing.Tape"
Wing.Tape
       -        +
2.134375 1.721250
```

The maximum flight time would be achieved with Paper.Type and Wing.Tape set low (light, no). The expected flight time with this combination is:

$$\begin{aligned} E[Y] &= 1.9278 - \frac{0.5006}{2}x_{\text{P}} - \frac{0.4131}{2}x_{\text{M}} \\ &= 1.9278 - \frac{0.5006}{2}(-1) - \frac{0.4131}{2}(-1) \\ &= 1.9278 + \frac{0.5006 + 0.4131}{2} \\ &= 2.38 \text{ sec} \end{aligned}$$

VIII.6 In a study to investigate several factors in the system of aircraft control a computer simulation model had to be used because of the legal and ethical problems with experimenting with an actual aircraft control system. This simulation model had been evolved over many years and had been verified using actual data. It is quite a complicated model in which many factors affected the final response, the time a pilot had to wait to speak to the controller; random variation was incorporated into the model.

It was desired to use the model to determine which factors affect the response and it was decided 8 factors would be investigated. The factors included the number of lengths of tracks within the sector, the number of adjacent high-altitude sectors, the mix of jumbo versus standard jets, and so on.

A full 2^8 design was impossible given the computer time required for each individual simulation. Instead it was decided to utilize a 2^{8-4} fraction with generators **I** = **1235**, **I** = **1246**, **I** = **1347** and **I** = **2348**. The results are given in the following table:

	Factor								
Simulation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	Time
1	-	-	-	-	-	-	-	-	65.81
2	+	-	-	-	+	+	+	-	58.49
3	-	+	-	-	+	+	-	+	62.51
4	+	+	-	-	-	-	+	+	60.19
5	-	-	+	-	+	-	+	+	60.22
6	+	-	+	-	-	+	-	+	59.20
7	-	+	+	-	-	+	+	-	66.58
8	+	+	+	-	+	-	-	-	61.68
9	-	-	-	+	-	+	+	+	59.01
10	+	-	-	+	+	-	-	+	53.71
11	-	+	-	+	+	-	+	-	62.43
12	+	+	-	+	-	+	-	-	60.77
13	-	-	+	+	+	+	-	-	60.44
14	+	-	+	+	-	-	+	-	57.48
15	-	+	+	+	-	-	-	+	63.08
16	+	+	+	+	+	+	+	+	58.32

Note that the aliasing pattern (ignoring three- and more-factor interactions) is as follows:

l_1 -> average
l_2 -> **1**
l_3 -> **2**
l_4 -> **12** + **35** + **46** + **78**
l_5 -> **3**
l_6 -> **13** + **25** + **47** + **68**
l_7 -> **23** + **15** + **48** + **67**
l_8 -> **5**
l_9 -> **4**
l_{10} -> **14** + **26** + **37** + **58**
l_{11} -> **24** + **16** + **38** + **57**
l_{12} -> **6**
l_{13} -> **34** + **17** + **28** + **56**
l_{14} -> **7**
l_{15} -> **8**
l_{16} -> **45** + **36** + **27** + **18**

What is the experimental structure for this experiment?

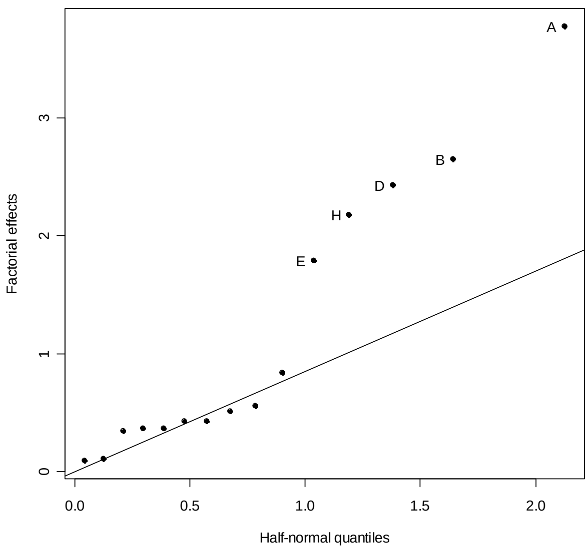
Structure	Formula
unrandomized	*16* Simulations
randomized	*2* A**2* B**2* C**2* D**2* E**2* F**2* G**2* H

On the basis of previous simulation studies it could be assumed that the standard deviation of an estimated effect was 0.35.

Analyze this data using R. Perform appropriate diagnostic checking.

```
> #
> # set up data frame
> #
> mp <- c("-", "+")
> fnames <- list(A = mp, B = mp, C = mp, D = mp)
> Frf8SimC.Treats <- fac.gen(generate = fnames, order = "yates")
> attach(Frf8SimC.Treats)
> Frf8SimC.Treats$E <- factor(mpone(A)*mpone(B)*mpone(C), labels = mp)
> Frf8SimC.Treats$FF <- factor(mpone(A)*mpone(B)*mpone(D), labels = mp)
> Frf8SimC.Treats$G <- factor(mpone(A)*mpone(C)*mpone(D), labels = mp)
> Frf8SimC.Treats$H <- factor(mpone(B)*mpone(C)*mpone(D), labels = mp)
> detach(Frf8SimC.Treats)
> Frf8SimC.dat <- data.frame(Runs = factor(1:16), Frf8SimC.Treats)
> remove("Frf8SimC.Treats")
> Frf8SimC.dat$Time <- c(65.81,58.49,62.51,60.19,60.22,59.20,66.58,61.68,59.01,
+                        53.71,62.43,60.77, 60.44,57.48,63.08,58.32)
> Frf8SimC.dat
   Runs A B C D E FF G H  Time
1     1 - - - - -  - - - 65.81
2     2 + - - - +  + + - 58.49
3     3 - + - - +  + - + 62.51
4     4 + + - - -  - + + 60.19
5     5 - - + - +  - + + 60.22
6     6 + - + - -  + - + 59.20
7     7 - + + - -  + + - 66.58
8     8 + + + - +  - - - 61.68
```

```
9      9 - - - + -  + + + 59.01
10    10 + - - + +  - - + 53.71
11    11 - + - + +  - + - 62.43
12    12 + + - + -  + - - 60.77
13    13 - - + + +  + - - 60.44
14    14 + - + + -  - + - 57.48
15    15 - + + + -  - - + 63.08
16    16 + + + + +  + + + 58.32
> #
> # analyse
> #
> Frf8SimC.aov <- aov(Time ~ (A + B + C + D + E + FF + G + H)^2 + Error(Runs),
Frf8SimC.dat)
> summary(Frf8SimC.aov)

Error: Runs
     Df Sum Sq Mean Sq
A     1 57.154  57.154
B     1 28.090  28.090
C     1  1.040   1.040
D     1 23.620  23.620
E     1 12.816  12.816
FF    1  0.032   0.032
G     1  1.254   1.254
H     1 19.010  19.010
A:B   1  0.548   0.548
A:C   1  0.548   0.548
A:D   1  0.048   0.048
A:E   1  0.740   0.740
A:FF  1  2.822   2.822
A:G   1  0.462   0.462
A:H   1  0.740   0.740
> qqyeffects(Frf8SimC.aov, error.term = "Runs", data=Frf8SimC.dat)
Effect(s) labelled: E H D B A
> round(yates.effects(Frf8SimC.aov, error.term="Runs", data=Frf8SimC.dat), 2)
    A     B     C     D     E    FF     G     H   A:B   A:C   A:D   A:E
-3.78  2.65  0.51 -2.43 -1.79  0.09 -0.56 -2.18  0.37  0.37  0.11  0.43
A:FF   A:G   A:H
0.84  0.34  0.43
```



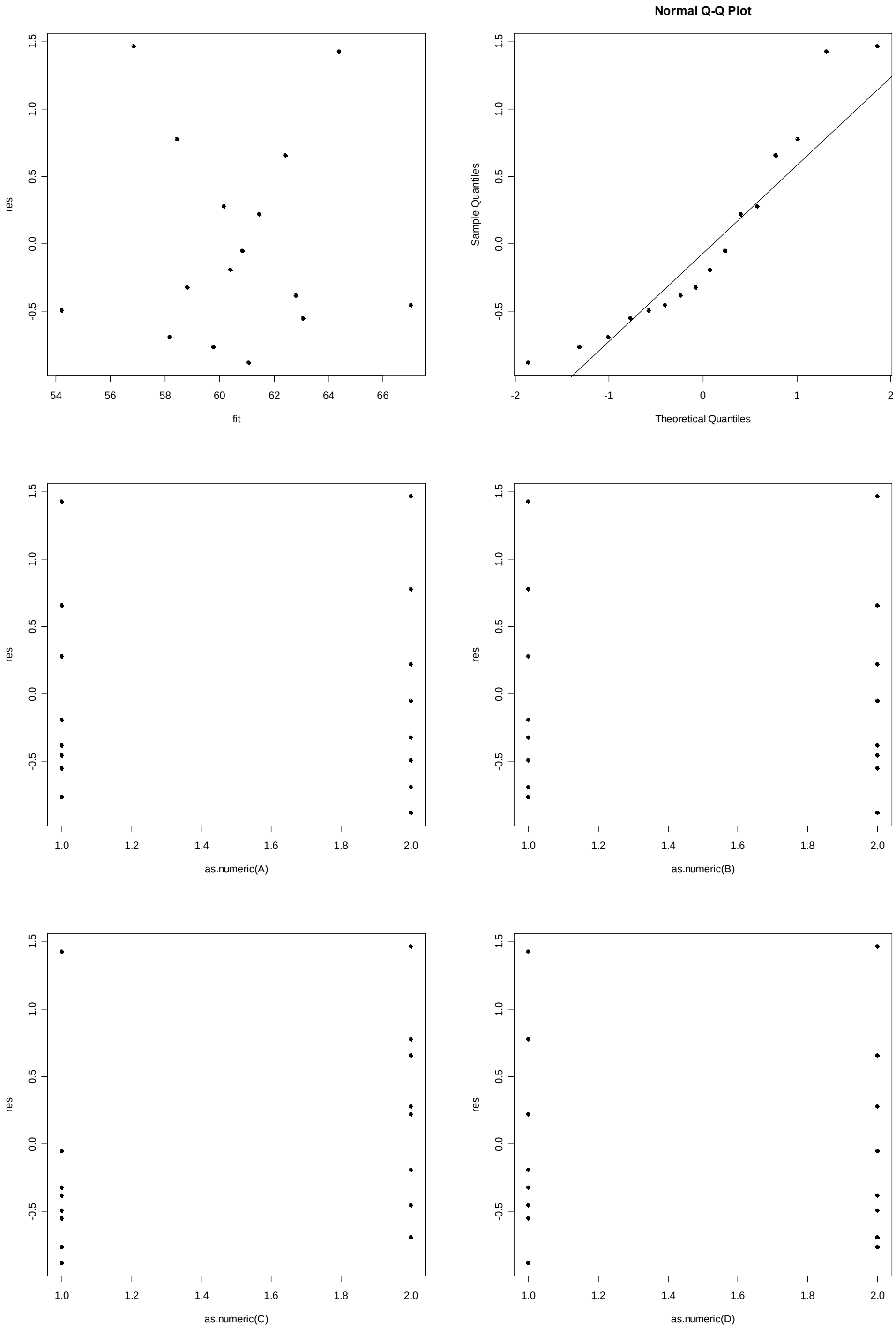
*The normal plot of Yates effects indicates that the significant effects from this analysis would appear to be **1**, **2**, **4**, **5** and **8**. The standard error of 0.35 would indicate that the interaction effect **16** is significant. However, the analysis of variance with this effect included indicates that the interaction effect is not significant. Thus the fitted model would appear to be*

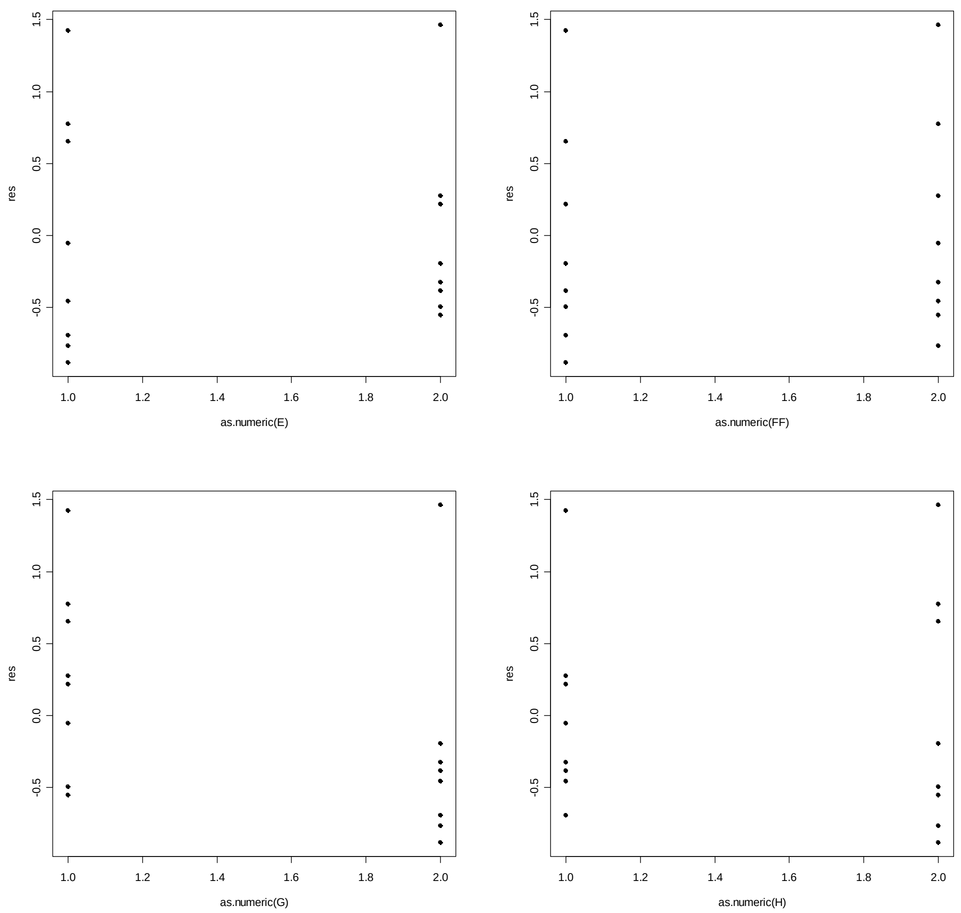
$$\psi = E[Y] = A + B + D + E + H.$$

```
> Frf8SimC.Fit.aov <- aov(Time ~ A*FF + B + D + E + H + Error(Runs),
+                                                        Frf8SimC.dat)
> summary(Frf8SimC.Fit.aov)

Error: Runs
          Df Sum Sq Mean Sq F value    Pr(>F)
A          1 57.154  57.154 84.9868 1.553e-05
FF         1  0.032   0.032  0.0482 0.8317607
B          1 28.090  28.090 41.7695 0.0001956
D          1 23.620  23.620 35.1221 0.0003512
E          1 12.816  12.816 19.0578 0.0023948
H          1 19.010  19.010 28.2671 0.0007139
A:FF       1  2.822   2.822  4.1969 0.0746670
Residuals  8  5.380   0.673
```

Because of the nonsignificance of A#FF we drop it from the model, along with FF. Apparently, the lack of effect of factor 6 had not been anticipated and caused the simulation model to be questioned and further runs to check this.

```
> #
> # Diagnostic checking
> #
> tukey.1df(Frf8SimC.Fit.aov, data=Frf8SimC.dat, error.term="Runs")
$Tukey.SS
[1] 0.003747143

$Tukey.F
[1] 0.004097202

$Tukey.p
[1] 0.9503619

$Devn.SS
[1] 8.231053

> res <- resid.errors(Frf8SimC.Fit.aov)
> fit <- fitted.errors(Frf8SimC.Fit.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> attach(Frf8SimC.dat)
> plot(as.numeric(A), res, pch=16)
> plot(as.numeric(B), res, pch=16)
> plot(as.numeric(C), res, pch=16)
> plot(as.numeric(D), res, pch=16)
> plot(as.numeric(E), res, pch=16)
> plot(as.numeric(FF), res, pch=16)
> plot(as.numeric(G), res, pch=16)
> plot(as.numeric(H), res, pch=16)
```


Normal Q-Q Plot
res
fit
Sample Quantiles
Theoretical Quantiles
as.numeric(A)
as.numeric(B)
as.numeric(C)
as.numeric(D)



The Tukey's one-degree-of-freedom-for-nonadditivity is not significant and the residual-versus-fitted-values and normal probability plots appear satisfactory. However, the residuals-versus C and G would seem to indicate differences in variance. This requires further investigation before the analysis can be accepted.