DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

X. Factorial designs at two levels

X.A Replicated 2^k experiments X-2
- a) Calculation of responses and Yates effects X-3
- b) Yates algorithm X-7
- c) Computation in Genstat X-7
- d) Treatment differences X-12

X.B Economy in experimentation X-15
- a) Initial analysis of variance X-16
- b) Analysis assuming no 3-factor or 4-factor interactions X-18
- c) Probability plot of Yates effects X-20
- d) Fitted values X-21
- e) Diagnostic checking X-22
- f) Computation in Genstat X-24

X.C Confounding in factorial experiments X-25
- a) Total confounding of effects X-25
- b) Partial confounding of effects X-32
- c) Computation in Genstat X-36

X.D Fractional factorial designs at two levels X-37
- a) Half-fractions of full factorial experiments X-44
- b) More on construction and use of half-factions X-52
- c) The concept of design resolution X-53
- d) Resolution III designs (also called main effect plans) X-54
- e) Table of fractional factorial designs X-62
- f) Generation of fractional factorial designs in Genstat X-63
- g) Computation in Genstat X-63

In this chapter we discuss the design and analysis of factorial experiments in which all factors have 2 levels.

Definition X.1: An experiment that involves *k* factors all at 2 levels is called a $\mathbf{2^k}$ **experiment**. ■

These designs represent an important class of designs for the following reasons:

1. They require relatively few runs per factor studied, and although they are unable to explore fully a wide region of the factor space, they can indicate trends and so determine a promising direction for further experimentation.

2. They can be suitably augmented to enable a more thorough local exploration.

3. They can be easily modified to form fractional designs in which only some of the treatment combinations are observed.

4. Their analysis and interpretation is relatively straightforward.

X.A Replicated 2^k experiments

(Box, Hunter and Hunter, ch. 10; Cochran & Cox, sec. 5.24–26, Mead, sec. 13.1–3, Mead & Curnow, sec. 6.6)

The design of this type of experiment is exactly the same as for the factorial experiments in general, as outlined in IX.A, *Design of factorial experiments*. However, the analysis is somewhat simpler. An experiment involving three factors — a 2^3 experiment — will be used to illustrate.

Definition X.2: There are three **systems of specifying treatment combinations** in common usage. The first uses a minus sign for the low level of a quantitative factor and a plus for the high level. Qualitative factors are coded arbitrarily but consistently as minus and plus. In the second notation, the upper level of a factor is denoted by a lower case letter used for that factor and the lower level by the absence of this letter. The third notation uses 0 and 1 in place of – and +. ■

We shall use the ± notation as it relates to the computations for the designs.

Example X.1 2^3 pilot plant experiment

An experimenter conducted a 2^3 experiment in which there are two quantitative factors — temperature and concentration — and a single qualitative factor — catalyst. Altogether 16 tests were conducted with the three factors assigned at random so that each occurred just twice. At each test the chemical yield was measured and the data is shown in the following table:

							Replicate	
T	C	K		T	C	K	1	2
–	–	–	1	0	0	0	59	61
+	–	–	t	1	0	0	74	70
–	+	–	c	0	1	0	50	58
+	+	–	tc	1	1	0	69	67
–	–	+	k	0	0	1	50	54
+	–	+	tk	1	0	1	81	85
–	+	+	ck	0	1	1	46	44
+	+	+	tck	1	1	1	79	81

This table also gives the treatment combinations, using the three systems, for the experiment and associated with each observation.

The components of this experiment are:

1. Observational unit – a test
2. Response variable – Yield
3. Unrandomized factors – Tests
4. Randomized factors – Temp, Conc, Catal
5. Type of study – Three-factor CRD

The experimental structure for this experiment is:

Structure	Formula
unrandomized	*16* Tests
randomized	*2* Temp**2* Conc**2* Catal

The terms derived from the randomized structure formula are:

Temp*Conc*Catal = Temp + (Conc*Catal) + Temp.(Conc*Catal)
= Temp + Conc + Catal + Conc.Catal
+ Temp.Conc + Temp.Catal + Temp.Conc.Catal

Since all the factors have two levels, the number of levels minus one will in every case be one. Thus, using the cross product rule, the degrees of freedom for any term will be a product of ones and hence be one.

Given that the only random factor is Tests, the following are the symbolic expressions for the maximal expectation and variation models:

$$E[Y] = \text{Temp.Conc.Catal}$$
$$\text{var}[Y] = \text{Tests}$$

■

a) Calculation of responses and Yates effects

When all the factors in a factorial experiment are at 2 levels the analysis presented in section IX.1 simplifies greatly. This is because main effects, elements of $\mathbf{a}_e$, $\mathbf{b}_e$ and $\mathbf{c}_e$, which are of the form $\bar{y}_i - \bar{y}_.$, for I = 1,2, simplify to

$$\bar{y}_1 - \bar{y}_. = \bar{y}_1 - \frac{\bar{y}_1 + \bar{y}_2}{2} = \frac{\bar{y}_1}{2} - \frac{\bar{y}_2}{2} = \frac{\bar{y}_1 - \bar{y}_2}{2} \text{ and } \bar{y}_2 - \bar{y}_. = \frac{\bar{y}_2 - \bar{y}_1}{2} = -\left(\bar{y}_1 - \bar{y}_.\right)$$

This is in some ways obvious. It is merely saying that the distance of one of two means from their mean is half the difference between two means. Note there is really only one independent main effect and this is reflected in the fact that they have just one degree of freedom.

Also, two-factor interactions, elements of $\mathbf{a.b}_e$, $\mathbf{a.c}_e$ and $\mathbf{b.c}_e$, are of the form $\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$, for $i = 1,2$; $j = 1,2$. For $i = 1$, $j = 1$ this simplifies to

$$\begin{aligned}\bar{y}_{11}-\bar{y}_{1.}-\bar{y}_{.1}+\bar{y}_{..} &= \bar{y}_{11}-\frac{\bar{y}_{11}+\bar{y}_{12}}{2}-\frac{\bar{y}_{11}+\bar{y}_{21}}{2}+\frac{\bar{y}_{11}+\bar{y}_{12}+\bar{y}_{21}+\bar{y}_{22}}{4}\\ &= \frac{4\bar{y}_{11}-2\bar{y}_{11}-2\bar{y}_{12}-2\bar{y}_{11}-2\bar{y}_{21}+\bar{y}_{11}+\bar{y}_{12}+\bar{y}_{21}+\bar{y}_{22}}{4}\\ &= \frac{\bar{y}_{11}-\bar{y}_{12}-\bar{y}_{21}+\bar{y}_{22}}{4}\end{aligned}$$

Similarly,

$$\begin{aligned}\bar{y}_{12}-\bar{y}_{1.}-\bar{y}_{.2}+\bar{y}_{..} &= \frac{-\bar{y}_{11}+\bar{y}_{12}+\bar{y}_{21}-\bar{y}_{22}}{4}\\ \bar{y}_{21}-\bar{y}_{2.}-\bar{y}_{.1}+\bar{y}_{..} &= \frac{-\bar{y}_{11}+\bar{y}_{12}+\bar{y}_{21}-\bar{y}_{22}}{4}\\ \bar{y}_{22}-\bar{y}_{2.}-\bar{y}_{.2}+\bar{y}_{..} &= \frac{\bar{y}_{11}-\bar{y}_{12}-\bar{y}_{21}+\bar{y}_{22}}{4}\end{aligned}$$

Once again there is only one independent quantity and so only one degree of freedom. Notice that what is computed is the difference between $\bar{y}_{12}-\bar{y}_{11}$ and $\bar{y}_{21}-\bar{y}_{22}$. Each of these is the difference between the first and second level of factor B, the simple effects of B, one for the first level and the other for the second level of A.

Indeed all effects in a 2^k experiment have only one degree of freedom, as contemplation of the cross-product rule for crossed factors will reveal. So to accomplish an analysis we actually only need to compute a single value for each effect, instead of a vector of effects. We do not compute exactly the quantities above. However, we do compute quantities that are proportional to them. We compute what are called the responses and from these the Yates main and interaction effects are computed.

Definition X.3: A **one-factor response** and a **Yates main effect** is the difference between the means for the high and low levels of the factor: $\bar{y}_{+}-\bar{y}_{-}$. ■

Definition X.4: A **two-factor response** is the difference of the simple effects: $(\bar{y}_{++}-\bar{y}_{+-})-(\bar{y}_{-+}-\bar{y}_{--})$. A **two-factor Yates interaction effect** is half the two-factor response. ■

Definition X.5: A **three-factor response** is the difference in the response of two factors at each level of the other factor:

$$(\bar{y}_{+++}-\bar{y}_{++-}-\bar{y}_{+-+}+\bar{y}_{+--})-(\bar{y}_{-++}-\bar{y}_{-+-}-\bar{y}_{--+}+\bar{y}_{---}).$$

A **three-factor Yates interaction** is the half difference in the Yates interaction effects of two factors at each level of the other factor; it is thus one-quarter of the response. ■

Definition X.6: **Sums of squares** can be computed from the Yates effects by squaring them and dividing by $2^k/r$ where k is the number of factors and r is the number of replicates of each treatment combination. ■

Example X.1 2^3 pilot plant experiment (continued)

We obtain the responses and Yates interaction effects for this experiment. This can be conveniently carried out on the means over the replicates.

The treatment means for the example are:

			Yield		
T	C	K	Rep 1	Rep 2	Mean
–	–	–	59	61	60
+	–	–	74	70	72
–	+	–	50	58	54
+	+	–	69	67	68
–	–	+	50	54	52
+	–	+	81	85	83
–	+	+	46	44	45
+	+	+	79	81	80

Thus, for the three factors in this experiment we have the following one-factor responses/main effects:

Temperature,

$$\frac{72+68+83+80}{4} - \frac{60+54+52+45}{4} = 75.75 - 52.75 = 23$$

Concentration,

$$\frac{54+68+45+80}{4} - \frac{60+72+52+83}{4} = 61.75 - 66.75 = -5$$

Catalyst,

$$\frac{52+83+45+80}{4} - \frac{60+72+54+68}{4} = 65.0 - 63.5 = 1.5$$

The two-factor T.K response is the difference between the simple effects of K for each T or the difference between the simple effects of T for each K — it does not matter which. The T.K Yates interaction effect is then half this response.

The simple effect of K

$$\text{at + (180°C)} = \frac{83+80}{2} - \frac{72+68}{2} = 81.5\text{-}70\text{=}11.5$$

$$\text{at – (160°C)} = \frac{52+45}{2} - \frac{60+54}{2} = 48.5\text{-}57\text{=-}8.5$$

so that the response is

$$11.5 - (-8.5) = 20$$

and the Yates interaction effect is 10.

The computation for the Yates interaction effect can be gathered together as follows:

$$\tfrac{1}{2}\left\{\frac{83+80}{2}-\frac{72+58}{2}-\frac{52+45}{2}+\frac{60+54}{2}\right\}$$

which is equal to

$$\frac{83+80+60+54}{4}-\frac{72+58+52+45}{4}$$

Thus, the Yates interaction is just the difference between two averages of four (half) of the observations. Similar results can be demonstrated for the other two two-factor interactions, T.C and C.K.

The three-factor T.C.K response is the half difference between the T.C interaction effects at each level of K. If one follows through the computations, in the same way as for two factor interaction above, it will be found that the three-factor Yates interaction effect consists of the difference between the following two means of four observations each:

$$\frac{72+54+52+80}{4}-\frac{60+68+83+45}{4}=64.5\text{-}64\text{=}0.5$$

Since, for the example, k=3 and r =2, the divisor for the sums of squares is $2^k/r = 2^3/2 = 4$. Hence, the T.C.K sums of squares is:

$$\text{T.C.K SSq} = \frac{0.5^2}{4} = 0.0625$$

■

It turns out that there are easy rules for determining the signs of observations to compute the Yates effects.

Definition X.7: The **signs for observations in a Yates effect** are obtained from the columns of pluses and minuses that specify the factor combinations for each observation by taking the columns for the factors in the effect and forming their elementwise product. The **elementwise product** is the result of multiplying pairs of elements in the same row as if they were ±1 and expressing the result as a ±. ■

Example X.1 2^3 pilot plant experiment (continued)

The table of coefficients is:

T	C	K	TC	TK	CK	TCK	Mean
–	–	–	+	+	+	–	60
+	–	–	–	–	+	+	72
–	+	–	–	+	–	+	54
+	+	–	+	–	–	–	68
–	–	+	+	–	–	+	52
+	–	+	–	+	–	–	83
–	+	+	–	–	+	–	45
+	+	+	+	+	+	+	80

Clearly, these can be used to assist in calculating the responses and effects and hence the sums of squares. ■

b) Yates algorithm

A particularly nifty method of performing the calculations is an algorithm due to Yates. It requires that the observations be in standard order. They are in standard order when the one column of plus/minuses consists of successive minus and plus signs, a second column of successive pairs of plus and minus signs, the third column of successive quadruplets of plus and minus signs and so forth. Yates algorithm for the example is given in the following table. Beginning with the means, one forms the sum of each successive pair of observations and the successive differences. This operation is repeated recursively until it has been performed k times. The Yates effects described above are then calculated by dividing all except the first by 2^{k-1}; the first is divided by 2^k. The effect can be identified by the pattern of pluses in its row. Thus, the 5th row contains Yates effect K which is equal to 1.5.

T	C	K	Mean	(1)	(2)	(3)	Estimate
–	–	–	60	132	254	514	64.25
+	–	–	72	122	260	92	23.0
–	+	–	54	135	26	–20	–5.0
+	+	–	68	125	66	6	1.5
–	–	+	52	12	–10	6	1.5
+	–	+	83	14	–10	40	10.0
–	+	+	45	31	2	0	0.0
+	+	+	80	35	4	2	0.5

c) Computation in Genstat

The analysis of replicated 2^k factorial experiments is the same as for the general factorial experiment. A small difference is the output of the table of Yates effects accomplished by specifying E in the PRINT option and Yates in the TWOLEVEL option.

Example X.1 2^3 pilot plant experiment (continued)

The Genstat output file containing the analysis of the example is:

```
Genstat 5  Release 4.1  (PC/Windows NT)              10 April 2000 13:55:23
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

                 Genstat 5 Fourth Edition - (for Windows)
                 Genstat 5 Procedure Library Release PL11
                 ______________________________________________

Genstat 5  Release 4.1  (PC/Windows NT)              13 April 2000 13:35:49
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

                 Genstat 5 Fourth Edition - (for Windows)
                 Genstat 5 Procedure Library Release PL11
                 ______________________________________________

   3  "Data taken from File: D:/ANALYSES/LM/MULTIFAC/FAC3PIL.GSH"
   4  DELETE [redefine=yes] Tests,Temp,Conc,Catal,Yield
   5  FACTOR [modify=yes;nvalues=16;levels=16] Tests
   6  READ Tests; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Tests        16         0        16

   8  FACTOR [modify=yes;nvalues=16;levels=2] Temp
   9  READ Temp; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Temp        16         0         2

  11  FACTOR [modify=yes;nvalues=16;levels=2] Conc
  12  READ Conc; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Conc        16         0         2

  14  FACTOR [modify=yes;nvalues=16;levels=2] Catal
  15  READ Catal; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Catal        16         0         2

  17  VARIATE [nvalues=16] Yield
  18  READ Yield

    Identifier   Minimum      Mean   Maximum    Values   Missing
         Yield     44.00     64.25     85.00        16         0

  20
  21  PRINT Tests,Catal,Conc,Temp,Yield

       Tests       Catal        Conc        Temp       Yield
           1           1           1           1       59.00
           2           1           1           1       61.00
           3           1           1           2       74.00
           4           1           1           2       70.00
           5           1           2           1       50.00
           6           1           2           1       58.00
           7           1           2           2       69.00
           8           1           2           2       67.00
           9           2           1           1       50.00
          10           2           1           1       54.00
          11           2           1           2       81.00
          12           2           1           2       85.00
```

```
         13           2           2           1       46.00
         14           2           2           1       44.00
         15           2           2           2       79.00
         16           2           2           2       81.00

  22  BLOCK Tests
  23  TREAT Temp*Catal*Conc
  24  ANOVA [PRINT=A,I,M,E; TWOLEV=Yates; FPROB=Y; PSE=LSD] Yield

24................................................................................

***** Analysis of variance *****

Variate: Yield

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Tests stratum
Temp                       1   2116.000   2116.000  264.50  <.001
Catal                      1      9.000      9.000    1.12  0.320
Conc                       1    100.000    100.000   12.50  0.008
Temp.Catal                 1    400.000    400.000   50.00  <.001
Temp.Conc                  1      9.000      9.000    1.12  0.320
Catal.Conc                 1      0.000      0.000    0.00  1.000
Temp.Catal.Conc            1      1.000      1.000    0.13  0.733
Residual                   8     64.000      8.000
Total                     15   2699.000


***** Tables of effects *****

Variate: Yield

***** Tests stratum *****

Temp Y-effect              23.00   s.e. 1.414     rep. 8

Catal Y-effect              1.50   s.e. 1.414     rep. 8

Conc Y-effect              -5.00   s.e. 1.414     rep. 8

Temp.Catal Y-effect        10.00   s.e. 1.414     rep. 4

Temp.Conc Y-effect          1.50   s.e. 1.414     rep. 4

Catal.Conc Y-effect         0.00   s.e. 1.414     rep. 4

Temp.Catal.Conc Y-effect
                            0.50   s.e. 1.414     rep. 2

***** Tables of means *****

Variate: Yield

Grand mean  64.25

     Temp        1        2
         52.75    75.75

    Catal        1        2
         63.50    65.00

     Conc        1        2
         66.75    61.75

     Temp    Catal        1        2
        1             57.00    48.50
        2             70.00    81.50

     Temp     Conc        1        2
        1             56.00    49.50
        2             77.50    74.00
```

```
    Catal     Conc         1        2
        1             66.00    61.00
        2             67.50    62.50

             Catal         1                 2
     Temp     Conc         1        2        1        2
        1             60.00    54.00    52.00    45.00
        2             72.00    68.00    83.00    80.00

*** Least significant differences of means (5% level) ***

Table                 Temp       Catal        Conc        Temp
                                                         Catal
rep.                     8           8           8           4
d.f.                     8           8           8           8
l.s.d.               3.261       3.261       3.261       4.612

Table                 Temp       Catal        Temp
                      Conc        Conc       Catal
                                              Conc
rep.                     4           4           2
d.f.                     8           8           8
l.s.d.               4.612       4.612       6.522

  25  APLOT METHOD=fit,normal

             -
             I
             I
             I
         4.0 I                  *
r            I
e            I
s            I               *                        *             *
i            I
d            I      *                  *         *              *
u        0.0 I
a            I
l            I      *                  *         *              *
s            I               *                        *             *
             I
             I
        -4.0 I                  *
             -+---------+---------+---------+---------+---------+---------+---
           40.0      48.0      56.0      64.0      72.0      80.0      88.0

                                         fitted values

                                          Normal plot

             -
             I
         4.0 I                                                           *
r            I
e            I
s            I                                           *  *    *
i            I
d            I                               *  *  * *
u        0.0 I
a            I
l            I                     * *  *  *
s            I         *    *  *
             I
             I
        -4.0 I *
             -+---------+---------+---------+---------+---------+---------+---
           -1.8      -1.2      -0.6       0.0       0.6       1.2       1.8

                                   expected Normal quantiles
```

Note that I have not included the instructions for Tukey's one-degree-of-freedom-for-nonadditivity. This is because the design is a CRD and it would not be produced for this design. If you did include the instructions, you will find that no covariate line is presented in the analysis of variance.

The hypothesis test for this experiment is given below. Notice that simplified versions of the hypotheses have been presented, this simplification being possible because all levels are at 2 factors.

Step 1: Set up hypotheses

a) H_0: $\alpha_1 = \alpha_2$
   H_1: $\alpha_1 \neq \alpha_2$

b) H_0: $\beta_1 = \beta_2$
   H_1: $\beta_1 \neq \beta_2$

c) H_0: $(\alpha\beta)_{22} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{11} = 0$
   H_1: $(\alpha\beta)_{22} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{11} \neq 0$

   Note that for $a = b = 2$ the expression $(\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..}$ reduces to $(\alpha\beta)_{22} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{11}$ for all *i* and *j*.

d) H_0: $\gamma_1 = \gamma_2$
   H_1: $\gamma_1 \neq \gamma_2$

e) H_0: $(\alpha\gamma)_{22} - (\alpha\gamma)_{12} - (\alpha\gamma)_{21} + (\alpha\gamma)_{11} = 0$
   H_1: $(\alpha\gamma)_{22} - (\alpha\gamma)_{12} - (\alpha\gamma)_{21} + (\alpha\gamma)_{11} \neq 0$

f) H_0: $(\beta\gamma)_{22} - (\beta\gamma)_{12} - (\beta\gamma)_{21} + (\beta\gamma)_{11} = 0$
   H_1: $(\beta\gamma)_{22} - (\beta\gamma)_{12} - (\beta\gamma)_{21} + (\beta\gamma)_{11} \neq 0$

g) H_0: $((\alpha\beta\gamma)_{222} - (\alpha\beta\gamma)_{122} - (\alpha\beta\gamma)_{212} + (\alpha\beta\gamma)_{112})$
   $- ((\alpha\beta\gamma)_{221} - (\alpha\beta\gamma)_{121} - (\alpha\beta\gamma)_{211} + (\alpha\beta\gamma)_{111}) = 0$
   H_1: $((\alpha\beta\gamma)_{222} - (\alpha\beta\gamma)_{122} - (\alpha\beta\gamma)_{212} + (\alpha\beta\gamma)_{112})$
   $- ((\alpha\beta\gamma)_{221} - (\alpha\beta\gamma)_{121} - (\alpha\beta\gamma)_{211} + (\alpha\beta\gamma)_{111}) \neq 0$

Step 2: Calculate test statistics

The analysis of variance table for the three-factor factorial CRD is:

Source	df	SSq	MSq	F	Prob
Tests	15	2699.0			
T	1	2116.0	2116.0	264.5	<0.001
C	1	100.0	100.0	12.5	0.008
T.C	1	9.0	9.0	1.1	0.320
K	1	9.0	9.0	1.1	0.320
T.K	1	400.0	400.0	50.0	<0.001
C.K	1	0.0	0.0	0.0	1.000
T.C.K	1	1.0	1.0	0.1	0.733
Residual	8	64.0	8.0		

Step 3: Decide between hypotheses

For T.C.K interaction: the T.C.K interaction is not significant.

For T.C, T.K and C.K interactions: only the T.K interaction is significant.

For C: the C effect is significant.

Thus, the yield depends on the particular combination of Temperature and Catalyst, whereas Concentration also affects the yield but independently of the other factors. Hence the model that best fits the data is:

$$E[Y] = C + T.K$$

d) Treatment differences

Mean differences

To examine how the factors affect the response one examines the tables of means corresponding to the terms in the fitted model. For the example, we examine the T.K and C tables of means.

The table of means for the T.K combinations is:

```
Temp     Catal         1        2
   1               57.00    48.50
   2               70.00    81.50
```

$$\begin{aligned} LSD(5\%) &= 2.306\sqrt{\frac{8\times 2}{4}} \\ &= 2.306\times 2.000 \\ &= 4.612 \end{aligned}$$

Thus, whereas the catalyst decreases yield at the lower temperature, it increases it at the higher temperature — a difference in direction interaction.

The table of Concentration means is

```
Conc        1        2
        66.75    61.75
```

It is evident that the higher concentration decreases the yield by about 5 units.

So what combination of treatments would give the highest yield? *Answer:* temperature and catalyst at the higher level and concentration at the lower level.

Polynomial models and fitted values

Given that there are only two levels of each factor a linear trend would fit perfectly the means of each factor. Now could fit linear trend models by putting the values of the factor levels for each factor as a column in an **X** matrix; a linear interaction term could be fitted by adding to the **X** matrix a column obtained by pairwise multiplying the values in the columns for the factors involved in the interaction. However, suppose that instead of putting the actual values of the factor levels into the **X** matrix, it was decided to code the values as ±1. Interaction terms can still be fitted as the pairwise products of the (coded) elements from the columns for the factors involved in the interaction. Now whether your fit is based on an **X** matrix with 0,1 or ±1s or the actual factor values, you end up with equivalent fits as the fitted values and F test statistics will be the same for all three parametrizations. The values of the parameter estimates will differ and you will need to put in the values you used in the **X** matrix to obtain the estimates. The advantage of using ±1 is the ease of obtaining the **X** matrix and the simplicity of the computations.

The columns of an **X** for a particular model can be obtained from the table of coefficients given for working out the sums of squares, with a column added for the grand mean term. You must include columns corresponding to each significant terms and any term marginal to a significant effect. The full table of coefficients is as follows:

I	T	C	K	TC	TK	CK	TCK
+	–	–	–	+	+	+	–
+	+	–	–	–	–	+	+
+	–	+	–	–	+	–	+
+	+	+	–	+	–	–	–
+	–	–	+	+	–	–	+
+	+	–	+	–	+	–	–
+	–	+	+	–	–	+	–
+	+	+	+	+	+	+	+

For the example, the significant terms are C and T.K so that the **X** matrix corresponding to this model would include columns for *I*, T, C, K and TK and the row for each treatment combination would be repeated *r* times. Thus, the linear trend model that best describes the data from the experiment is:

$$E[\mathbf{Y}] = \mathbf{X}\theta = \left(\begin{bmatrix} 1 & -1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \otimes \mathbf{1}_r \right) \begin{bmatrix} \mu \\ \gamma_{\mathrm{T}} \\ \gamma_{\mathrm{C}} \\ \gamma_{\mathrm{K}} \\ \gamma_{\mathrm{TK}} \end{bmatrix}$$

We can write an element of $E[\mathbf{Y}]$ as

$$E[Y_k] = \mu + \gamma_{\mathrm{T}} x_{\mathrm{T}} + \gamma_{\mathrm{C}} x_{\mathrm{C}} + \gamma_{\mathrm{K}} x_{\mathrm{K}} + \gamma_{\mathrm{TK}} x_{\mathrm{T}} x_{\mathrm{K}}$$

where x_{T}, x_{C} and x_{K} takes values ±1 according to whether the observation took the high or low level of the factor.

Theorem X.1: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ with $\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ where the elements of **X** are ±1 and, for $\mathbf{x}_i$ and $\mathbf{x}_j$ columns of **X**,

$$\mathbf{x}_i'\mathbf{x}_j = \begin{cases} n & i = j \\ 0 & i \neq j \end{cases} \text{ with } n = 2^k r$$

Then, $\hat{\theta}_i = n^{-1}\mathbf{x}_i'\mathbf{Y}$.

Proof: Left as an exercise ■

Now it is clear that the columns in the tables of coefficients used to compute the Yates effect conform to the conditions placed on the columns of **X** in this theorem. Further, it can be seen that the estimator of a Yates effect is $2n^{-1}\mathbf{x}_i'\mathbf{Y}$. Hence, an estimator of one of coefficients in the model is half a Yates effect, with the estimator for the first column being the grand mean.

For the example, the fitted model for a single observation is thus

$$E[Y_k] = 64.25 + \frac{23.0}{2} x_{\mathrm{T}} + \frac{1.5}{2} x_{\mathrm{K}} + \frac{20.0}{2} x_{\mathrm{T}} x_{\mathrm{K}} - \frac{5.0}{2} x_{\mathrm{C}}$$

The optimum yield occurs for T and K high and C low so it is estimated to be

$$\begin{aligned} E[Y_k] &= 64.25 + 11.5 \times 1 + 0.75 \times 1 + 10 \times 1 \times 1 - 2.5 \times (-1) \\ &= 89 \end{aligned}$$

Also note that a particular table of means can be obtained by using a linear trend model that includes the term corresponding to the table of means and any marginal teams. Hence, the table of T.K means can be obtained by substituting $x_{\mathrm{T}} = \pm 1$, $x_{\mathrm{K}} = \pm 1$ into

$$E[Y_k] = 64.25 + \frac{23.0}{2} x_{\mathrm{T}} + \frac{1.5}{2} x_{\mathrm{K}} + \frac{20.0}{2} x_{\mathrm{T}} x_{\mathrm{K}}$$

X.B Economy in experimentation

(Box, Hunter and Hunter, sec.10.8)

In most situations there are more factors to be investigated than can be conveniently accommodated with the time and budget available. Rather than duplicate a 2^3 factorial as was done in the pilot plant study, it is usually better to include a fourth variable and run an unreplicated 2^4 design. Or, as we shall see in later sections, run a half-replicated 2^5 and use 16 runs to investigate 5 factors.

The problem that would appear to arise from this suggestion is that if there is no replication then it will be impossible to measure the uncontrolled variation that has occurred in the experiment. However, when there are 4 or more factors it is unlikely that all factors will affect the response. Further it is usual that the magnitudes of effects are getting smaller as the order of the effect increases. Thus, it is likely that three-factor and higher-order interactions will be small and can be ignored without seriously affecting the conclusions drawn from the experiment.

Example X.2 A 2^4 process development study

The data given in the table below are the results, taken from Box, Hunter and Hunter, from a 2^4 design employed in a process development study.

K	T	P	C	Conversion (%)	Order of Runs
–	–	–	–	71	(8)
+	–	–	–	61	(2)
–	+	–	–	90	(10)
+	+	–	–	82	(4)
–	–	+	–	68	(15)
+	–	+	–	61	(9)
–	+	+	–	87	(1)
+	+	+	–	80	(13)
–	–	–	+	61	(16)
+	–	–	+	50	(5)
–	+	–	+	89	(11)
+	+	–	+	83	(14)
–	–	+	+	59	(3)
+	–	+	+	51	(12)
–	+	+	+	85	(6)
+	+	+	+	78	(7)

a) Initial analysis of variance

The following are the Genstat instructions for an initial analysis of this data:

```
PRINT Runs,Conc,Press,Temp,Catal,Conv
BLOCK Runs
TREAT Temp*Catal*Conc*Press
ANOVA [PRINT=A,I,E; FACTORIAL=4; TWOLEVEL=YATES; FPROB=Y] Conv
```

Note that the FACTORIAL option has to be set to 4 in the ANOVA command so that terms with up to 4 factors are included in the analysis. Otherwise, the default value of three will be used and the four-factor terms omitted.

The Genstat output file produced by these instructions is as follows:

```
Genstat 5  Release 4.1  (PC/Windows NT)              10 April 2000 14:39:54
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

                 Genstat 5 Fourth Edition - (for Windows)
                 Genstat 5 Procedure Library Release PL11
                 ______________________________________________

   3  "Data taken from File: D:/ANALYSES/LM/MULTIFAC/FAC4PROC.GSH"
   4  DELETE [redefine=yes] Runs,Conc,Press,Temp,Catal,Conv
   5  FACTOR [modify=yes;nvalues=16;levels=16] Runs
   6  READ Runs; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Runs        16         0        16

   8  FACTOR [modify=yes;nvalues=16;levels=2] Conc
   9  READ Conc; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Conc        16         0         2

  11  FACTOR [modify=yes;nvalues=16;levels=2] Press
  12  READ Press; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Press        16         0         2

  14  FACTOR [modify=yes;nvalues=16;levels=2] Temp
  15  READ Temp; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Temp        16         0         2

  17  FACTOR [modify=yes;nvalues=16;levels=2] Catal
  18  READ Catal; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Catal        16         0         2

  20  VARIATE [nvalues=16] Conv
  21  READ Conv

    Identifier   Minimum      Mean   Maximum    Values   Missing
          Conv     50.00     72.25     90.00        16         0

  23
```

```
  24  PRINT Runs,Conc,Press,Temp,Catal,Conv

        Runs        Conc       Press        Temp       Catal        Conv
           1           1           1           1           1       71.00
           2           1           1           1           2       61.00
           3           1           1           2           1       90.00
           4           1           1           2           2       82.00
           5           1           2           1           1       68.00
           6           1           2           1           2       61.00
           7           1           2           2           1       87.00
           8           1           2           2           2       80.00
           9           2           1           1           1       61.00
          10           2           1           1           2       50.00
          11           2           1           2           1       89.00
          12           2           1           2           2       83.00
          13           2           2           1           1       59.00
          14           2           2           1           2       51.00
          15           2           2           2           1       85.00
          16           2           2           2           2       78.00

  25  BLOCK Runs
  26  TREAT Temp*Catal*Conc*Press
  27  ANOVA [PRINT=A,I,E; FACTORIAL=4; TWOLEVEL=YATES; FPROB=Y] Conv


27................................................................................

***** Analysis of variance *****

Variate: Conv

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Runs stratum
Temp                       1    2304.00    2304.00
Catal                      1     256.00     256.00
Conc                       1     121.00     121.00
Press                      1      20.25      20.25
Temp.Catal                 1       4.00       4.00
Temp.Conc                  1      81.00      81.00
Catal.Conc                 1       0.00       0.00
Temp.Press                 1       6.25       6.25
Catal.Press                1       2.25       2.25
Conc.Press                 1       0.25       0.25
Temp.Catal.Conc            1       1.00       1.00
Temp.Catal.Press           1       2.25       2.25
Temp.Conc.Press            1       2.25       2.25
Catal.Conc.Press           1       0.25       0.25
Temp.Catal.Conc.Press      1       0.25       0.25
Total                     15    2801.00


***** Tables of effects *****

Variate: Conv

***** Runs stratum *****

Temp Y-effect              24.00   s.e. *         rep. 8

Catal Y-effect             -8.00   s.e. *         rep. 8

Conc Y-effect              -5.50   s.e. *         rep. 8

Press Y-effect             -2.25   s.e. *         rep. 8

Temp.Catal Y-effect         1.00   s.e. *         rep. 4

Temp.Conc Y-effect          4.50   s.e. *         rep. 4

Catal.Conc Y-effect         0.00   s.e. *         rep. 4
```

```
Temp.Press Y-effect        -1.25   s.e. *         rep. 4

Catal.Press Y-effect        0.75   s.e. *         rep. 4

Conc.Press Y-effect        -0.25   s.e. *         rep. 4

Temp.Catal.Conc Y-effect
                            0.50   s.e. *         rep. 2

Temp.Catal.Press Y-effect
                           -0.75   s.e. *         rep. 2

Temp.Conc.Press Y-effect
                           -0.75   s.e. *         rep. 2

Catal.Conc.Press Y-effect
                           -0.25   s.e. *         rep. 2

Temp.Catal.Conc.Press Y-effect
                           -0.25   s.e. *         rep. 1
```

Now, as discussed before, there is no direct estimates of σ_R^2, the uncontrolled variation, available as there were no replicates in the 16 runs.

b) Analysis assuming no 3-factor or 4-factor interactions

However, if we assume that all three-factor and four-factor interactions are negligible, then we could use these to estimate the uncontrolled variation as this is the only reason for them being nonzero. To do this rerun the analysis with the FACTORIAL option set to 2. The output produced is as follows:

```
  28
  29  "Perform analysis assuming 3- & 4-factor interactions negligible"
  30  ANOVA [PRINT=A,I,E; FACTORIAL=2; TWOLEVEL=YATES; FPROB=Y] Conv


30.............................................................................

***** Analysis of variance *****

Variate: Conv

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Runs stratum
Temp                       1   2304.000   2304.000 1920.00  <.001
Catal                      1    256.000    256.000  213.33  <.001
Conc                       1    121.000    121.000  100.83  <.001
Press                      1     20.250     20.250   16.87  0.009
Temp.Catal                 1      4.000      4.000    3.33  0.127
Temp.Conc                  1     81.000     81.000   67.50  <.001
Catal.Conc                 1      0.000      0.000    0.00  1.000
Temp.Press                 1      6.250      6.250    5.21  0.071
Catal.Press                1      2.250      2.250    1.87  0.229
Conc.Press                 1      0.250      0.250    0.21  0.667
Residual                   5      6.000      1.200
Total                     15   2801.000


* MESSAGE: the following units have large residuals.

Runs 12           1.25   s.e. 0.61


***** Tables of effects *****

Variate: Conv
```

```
***** Runs stratum *****

Temp Y-effect              24.00   s.e. 0.548     rep. 8

Catal Y-effect             -8.00   s.e. 0.548     rep. 8

Conc Y-effect              -5.50   s.e. 0.548     rep. 8

Press Y-effect             -2.25   s.e. 0.548     rep. 8

Temp.Catal Y-effect         1.00   s.e. 0.548     rep. 4

Temp.Conc Y-effect          4.50   s.e. 0.548     rep. 4

Catal.Conc Y-effect         0.00   s.e. 0.548     rep. 4

Temp.Press Y-effect        -1.25   s.e. 0.548     rep. 4

Catal.Press Y-effect        0.75   s.e. 0.548     rep. 4

Conc.Press Y-effect        -0.25   s.e. 0.548     rep. 4
```

The analysis is summarized in the following analysis of variance table:

Source	df	SSq	MSq	F	Prob
Runs	15	2801.00			
Temp	1	2304.00	2304.00	1920.00	<0.001
Catal	1	256.00	256.00	213.33	<0.001
Conc	1	121.00	121.00	100.83	<0.001
Press	1	20.25	20.25	16.87	0.009
Temp.Catal	1	4.00	4.00	3.33	0.127
Temp.Conc	1	81.00	81.00	67.50	<0.001
Catal.Conc	1	0.00	0.00	0.00	1.000
Temp.Press	1	6.25	6.25	5.21	0.071
Catal.Press	1	2.25	2.25	1.87	0.229
Conc.Press	1	0.25	0.25	0.21	0.667
Residual	5	6.00	1.20		

This analysis indicates that there is an interaction between Temperature and Concentration and that Pressure and Catalyst also affect the Conversion percentage, although independently of the other factors.

However, there is a problem with this in that the test for main effects has been preceded by a test for interaction terms. Thus, testing is not independent and an allowance needs to be made for this. A further general problem in the use of higher-order interactions for error is that occasionally meaningful higher order interactions occur. The analysis presented above does not confront either of these problems.

c) Probability plot of Yates effects

A method that does not require the assumption of zero higher-order interactions and allows for the dependence of the testing is a Normal probability plot of the Yates effects. In this plot the Yates effects are plotted against standard normal deviates. This is done on the basis that if there were no effects of the factors, the estimated effects would be just normally distributed uncontrolled variation. Under these circumstances a straight-line plot of normal deviates versus effects is expected.

The procedure A2PLOT, preceded by BLOCK and TREAT directives, produces a normal plot for the Yates effects as shown in the following output:

```
  31
  32  "Produce normal plot of Yates effects"
  33  A2PLOT [PRINT=E; FACTORIAL=4; STRATUM=Runs; METHOD=normal; \
  34          GRAPH=line] Conv


34.................................................................................

***** Tables of effects *****

Variate: Conv

***** Runs stratum *****

Temp Y-effect              24.00   s.e. *        rep. 8

Catal Y-effect             -8.00   s.e. *        rep. 8

Conc Y-effect              -5.50   s.e. *        rep. 8

Press Y-effect             -2.25   s.e. *        rep. 8

Temp.Catal Y-effect         1.00   s.e. *        rep. 4

Temp.Conc Y-effect          4.50   s.e. *        rep. 4

Catal.Conc Y-effect         0.00   s.e. *        rep. 4

Temp.Press Y-effect        -1.25   s.e. *        rep. 4

Catal.Press Y-effect        0.75   s.e. *        rep. 4

Conc.Press Y-effect        -0.25   s.e. *        rep. 4

Temp.Catal.Conc Y-effect
                            0.50   s.e. *        rep. 2

Temp.Catal.Press Y-effect
                           -0.75   s.e. *        rep. 2

Temp.Conc.Press Y-effect
                           -0.75   s.e. *        rep. 2

Catal.Conc.Press Y-effect
                           -0.25   s.e. *        rep. 2

Temp.Catal.Conc.Press Y-effect
                           -0.25   s.e. *        rep. 1
```

```
             -
N            I
o            I                                                        *
r     1.3667 I                       *
m            I
a            I                **
l            I                *
             I               2
Q            I               *
u    -0.2733 I              **
a            I              *
n            I           * *
t            I
i            I      *
l            I *
e    -1.9133 I
             -+---------+---------+---------+---------+---------+---------+-
         -8.800    -2.933     2.933     8.800    14.667    20.533    26.400

                                            effect
```

Now we see that two negative effects and two positive effects deviate substantially from the straight line going through the remainder of the effects. Examination of the Yates effects reveals that the large negative effects correspond to Catalyst and Concentration and that the large positive effects correspond to Temperature and Temperature.Concentration. Hence, Temperature and Concentration interact in their affect on the Conversion percentage and Catalyst affects the response independently of any other factors.

d) Fitted values

Definition X.8: The **fitted values** are obtained using the **fitted equation** that consists of the grand mean, the terms for each significant effect and the effects marginal to them. A term consists of the product of x variables, one for each factor in the term; the x variables take the values –1 and +1 according whether the fitted value is required for the low or high level of that factor. The coefficient of the term is half the Yates interaction effect. ■

For the example, the fitted equation incorporating the significant effects is:

$$E[Y_k] = 72.25 - \frac{8.0}{2}x_{\mathrm{K}} + \frac{24.0}{2}x_{\mathrm{T}} - \frac{5.5}{2}x_{\mathrm{C}} + \frac{4.5}{2}x_{\mathrm{T}}x_{\mathrm{C}}$$

where x_{K}, x_{T} and x_{C} take the values –1 and +1.

To predict the response for a particular combination of the treatments, substitute in the appropriate combination of –1 and +1. For example, the predicted response for high catalyst and temperature but a low concentration is calculated as follows:

$$\begin{aligned} E[Y_k] &= 72.25 - \frac{8.0}{2}(+1) + \frac{24.0}{2}(+1) - \frac{5.5}{2}(-1) + \frac{4.5}{2}(+1)(-1) \\ &= 72.25 + \frac{-8.0 + 24.0 + 5.5 - 4.5}{2} = 80.75 \end{aligned}$$

e) Diagnostic checking

Having determined the significant terms one can reanalyze with just these terms, and those marginal to them, included in the TREAT structure and obtain the Residuals from this model. The Residuals can be used to do the usual diagnostic checking. For this to be effective requires that the number of fitted effects is small compared to the total number of effects in the experiment.

For the example the Genstat output file is as follows:

```
  35
  36  "Perform analysis including only significant effects
 -37     and do Residual analysis"
  38  BLOCK Runs
  39  TREAT Temp*Conc+Catal
  40  ANOVA [FPROB=Y; PSE=LSD] Conv


40..............................................................................

***** Analysis of variance *****

Variate: Conv

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Runs stratum
Temp                       1   2304.000   2304.000  649.85  <.001
Conc                       1    121.000    121.000   34.13  <.001
Temp.Conc                  1     81.000     81.000   22.85  <.001
Catal                      1    256.000    256.000   72.21  <.001
Residual                  11     39.000      3.545
Total                     15   2801.000


* MESSAGE: the following units have large residuals.

Runs 12           3.25   s.e. 1.56


***** Tables of means *****

Variate: Conv

Grand mean  72.25

     Temp       1        2
            60.25    84.25

     Conc       1        2
            75.00    69.50

     Temp     Conc        1        2
        1             65.25    55.25
        2             84.75    83.75

    Catal       1        2
            76.25    68.25
```

```
*** Least significant differences of means (5% level) ***

Table                  Temp        Conc        Temp       Catal
                                               Conc
rep.                      8           8           4           8
d.f.                     11          11          11          11
l.s.d.                2.072       2.072       2.930       2.072

  41  APLOT METHOD=fit,normal


             -
             I
             I
             I
         3.0 I                                        *
r            I
e            I
s            I              *            *
i            I                                         *        **
d            I
u        0.0 I
a            I    *         *  2
l            I                                         *
s            I    *                      *
             I                                        *          *
             I
        -3.0 I                                                  *
             -+---------+---------+---------+---------+---------+---------+---
           48.0      56.0      64.0      72.0      80.0      88.0      96.0

                                         fitted values


                                          Normal plot

             -
             I
         3.0 I                                                           *
r            I
e            I
s            I                                              *    *
i            I                                     * *   *
d            I
u        0.0 I
a            I                          *  * *  *
l            I                       *
s            I                 *   *
             I         *    *
             I
        -3.0 I *
             -+---------+---------+---------+---------+---------+---------+---
           -1.8      -1.2      -0.6       0.0       0.6       1.2       1.8

                                   expected Normal quantiles

  42  "
 -43  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -44  **** It is the term designated covariate in the following analysis
 -45  "
  46  AKEEP [FIT=Fit]
  47  CALC ResSq=Fit*Fit
  48  ANOVA [PRINT=*]  ResSq; RES=ResSq
  49  COVAR ResSq                               "A computational trick"
  50  ANOVA [PRINT=A; FPROB=Y] Conv
```

```
50.............................................................................

***** Analysis of variance (adjusted for covariate) *****

Variate: Conv
Covariate: ResSq

Source of variation     d.f.       s.s.       m.s.    v.r. cov.ef.  F pr.

Runs stratum
Temp                       1   2304.000   2304.000  665.89    1.00  <.001
Conc                       1    121.000    121.000   34.97    1.00  <.001
Temp.Conc                  1     81.000     81.000   23.41    1.00  <.001
Catal                      1    256.000    256.000   73.99    1.00  <.001
Covariate                  1      4.399      4.399    1.27          0.286
Residual                  10     34.601      3.460            1.02
Total                     15   2801.000

  51  COVAR
```

The residual-versus-fitted-values plot is displaying some curvature and heterogeneity of variance; however, Tukey's one-degree-of-freedom-for-nonadditivity is not significant. The Normal Probability plot shows a straight line trend. Some investigation of the need for transformation is required.

f) Computation in Genstat

The analysis of single replicate of a 2^k factorial experiments is based the normal probability plot of Yates effects, followed by an ANOVA for fitted model so that tables of means and residuals can be obtained and diagnostic checking performed.

Example X.2 A 2^4 process development study (continued)

The Genstat instructions to analyse this data are as follows:

```
PRINT Runs,Conc,Press,Temp,Catal,Conv
"Produce normal plot of Yates effects"
BLOCK Runs
TREAT Temp*Catal*Conc*Press
A2PLOT [PRINT=E; FACTORIAL=4; STRATUM=Runs; METHOD=normal; \
        GRAPH=line] Conv
"Perform analysis including only significant effects
   and do Residual analysis"
BLOCK Runs
TREAT Temp*Conc+Catal
ANOVA [FPROB=Y; PSE=LSD] Conv
APLOT METHOD=fit,normal
"
**** Tukey''s one-degree-of-freedom-for-non-additivity.
**** It is the term designated covariate in the following analysis
"
AKEEP [FIT=Fit]
CALC ResSq=Fit*Fit
ANOVA [PRINT=*]  ResSq; RES=ResSq
COVAR ResSq                             "A computational trick"
ANOVA [PRINT=A; FPROB=Y] Conv
COVAR
```

X.C Confounding in factorial experiments

a) Total confounding of effects

It happens it is not always possible to get a complete set of the treatments into a block or row of an experiment. The incomplete block designs and Youden square designs are available for experiments that that involve just a single set of treatments. However, the need for incomplete sets of treatments in a block is even more of a problem with factorial experiments where the number of treatments tends to be larger; on occasion, very large since the number of treatments increases geometrically with the number of factors and levels employed in the experiment. The solution to the problem must take into account the several factors. This turns out to be somewhat of an advantage in that it is possible to have incomplete sets of treatments and retain the orthogonality of the analysis.

Definition X.9: A **confounded factorial experiment** is one in which incomplete sets of treatments occur in each block.

The choice of which treatments to put in each block is done by deciding which effect is to be *confounded* with block differences.

Definition X.10: A **generator** for a confounded experiment is a relationship that specifies which effect is equal to a particular block contrast.

Example X.3 Complete sets of factorial treatments in 2 blocks

Suppose that a trial is to be conducted using a 2^3 factorial design and, to make the eight runs as homogeneous as possible, it desirable that batches of raw material sufficient for a complete set of treatments be blended together. However, suppose that the available blender can only blend sufficient for four runs at a time. This means that two blends will be required for a complete set of treatments.

It seems intuitively reasonable that we should arrange the treatments into two sets in such a way that the effect on the conclusions drawn from the experiment is minimized. Thus, the least serious thing to do is to have the three factor interaction mixed up or *confounded* with blocks and the other effects unconfounded. To arrange for this to happen is easily done with the table of pluses and minuses. The following is the table for the three factors:

Treatment	A	B	C	AB	AC	BC	ABC
1	–	–	–	+	+	+	–
2	+	–	–	–	–	+	+
3	–	+	–	–	+	–	+
4	+	+	–	+	–	–	–
5	–	–	+	+	–	–	+
6	+	–	+	–	+	–	–
7	–	+	+	–	–	+	–
8	+	+	+	+	+	+	+

Suppose we arrange to have all the treatments with a minus in the ABC column of the table together using one blend and the treatments with a plus in the ABC column using a second blend. Then we have divided the 8 treatments into 2 groups based on the minuses and pluses in the ABC column. In laying out the experiment the two groups are randomly assigned to the two different blends. That is, the allocation to blends might be as shown below:

Treatment	A	B	C	AB	AC	BC	ABC	Group	Blend
1	–	–	–	+	+	+	–	1	2
2	+	–	–	–	–	+	+	2	1
3	–	+	–	–	+	–	+	2	1
4	+	+	–	+	–	–	–	1	2
5	–	–	+	+	–	–	+	2	1
6	+	–	+	–	+	–	–	1	2
7	–	+	+	–	–	+	–	1	2
8	+	+	+	+	+	+	+	2	1

We see that the Blend difference has been associated, and hence confounded, with the ABC effect. The generator for this design is thus Blend = ABC. Examination of this table reveals that all other treatments have two minus and two plus observations in each blend. Hence, they are not affected by blend.

The analysis can still be accomplished using Yates algorithm on the 8 runs with the three treatment factors; it just has to be remembered that the A.B.C term is associated with Blend differences.

The experimental structure and the analysis of variance table for this experiment are:

Structure	Formula
Unrandomized	*2* Blends/*4* Runs
Randomized	*2* A**2* B**2* C

Source	df	E[MSq]
Blends	1	
A.B.C	1	$\sigma^2_{\text{BR}} + 2\sigma^2_{\text{B}} + f_{\text{A.B.C}}(\psi)$
Blends.Runs	6	
A	1	$\sigma^2_{\text{BR}} + f_{\text{A}}(\psi)$
B	1	$\sigma^2_{\text{BR}} + f_{B}(\psi)$
A.B	1	$\sigma^2_{\text{BR}} + f_{\text{A.B}}(\psi)$
C	1	$\sigma^2_{\text{BR}} + f_{\text{C}}(\psi)$
A.C	1	$\sigma^2_{\text{BR}} + f_{\text{A.C}}(\psi)$
B.C	1	$\sigma^2_{\text{BR}} + f_{\text{B.C}}(\psi)$

In this experiment, we have gained the advantage of having blocks of size 4 but at the price of being unable to estimate the three factor interaction. As can be seen from the expected mean squares, Blend variability and the A.B.C interaction cannot be estimated separately. This is not a problem if the interaction can be assumed to be negligible.

A layout for this design can be generated in Genstat using the *Stats > Design > Select Design* command. You need to select the *factorial designs (with interactions confounded with blocks)* option and the design option *Single replicate of a 2x2x2 factorial in blocks of size 4.* You will then be asked a series of questions to which you should respond as follows:

How many times would you like to replicate the basic design?	1
What would you like to call treatment factor 1?	name for factor 1
What would you like to call treatment factor 2?	name for factor 2
What would you like to call treatment factor 3?	name for factor 3
What would you like to call block factor 1?	name for block factor
What would you like to call block factor 2?	name for units
Seed for randomization (0 for none)?	6-digit number
Do you want to print the design?	yes
Do you want to check the design by ANOVA?	yes
Do you want to form a Units label variate?	no

Example X.4 Repeated two block experiment

It may be considered desirable to increase the size of the experiment to increase the precision with which the effects are estimated. If there are no extra factors available for inclusion, then the basic design could be replicated say r times which requires $2r$ blends. There is a choice as to how the 2 groups of treatments are to be assigned to the blends. For example, the groups of treatments could be assigned completely at random so that each group occurred with r out of the $2r$ blends. Another possibility is that the blends are formed into blocks of two and the groups of treatments randomized to the two blends within each block; to make this worthwhile would need to be able to identify relatively similar pairs of blends, otherwise complete randomization would be preferable. The experimental structure for the completely randomized case is as for the previous experiment, except that there would be $2r$ blends. The analysis would be:

Source	df	E[MSq]
Blends	$2r–1$	
A.B.C	1	$\sigma^2_{\text{BR}} \; +2\sigma^2_{\text{B}} \; +f_{\text{A.B.C}}(\psi)$
Residual	$2(r–1)$	$\sigma^2_{\text{BR}} \; +2\sigma^2_{\text{B}}$
Blends.Runs	$6r$	
A	1	$\sigma^2_{\text{BR}} \; +f_{\text{A}}(\psi)$
B	1	$\sigma^2_{\text{BR}} \; +f_{B}(\psi)$
A.B	1	$\sigma^2_{\text{BR}} \; +f_{\text{A.B}}(\psi)$
C	1	$\sigma^2_{\text{BR}} \; +f_{\text{C}}(\psi)$
A.C	1	$\sigma^2_{\text{BR}} \; +f_{\text{A.C}}(\psi)$
B.C	1	$\sigma^2_{\text{BR}} \; +f_{\text{B.C}}(\psi)$
Residual	$6(r–1)$	σ^2_{BR}

A layout for this design can be generated in Genstat using the *Stats > Design > Select Design* command as before except that you must specify the *number of replicates of the basic design* to be *r*. You will need to also supply the name of a factor to index the replicates, for example Reps.

Example X.5 Complete sets of factorial treatments in 4 blocks

Suppose that a 2^3 experiment is to be run but that the blends are only large enough for two runs using one blend. How can we design the experiment best? There will be four groups of treatments which we can represent using two factors at two levels. Let's suppose it is decided to associate the ABC interaction and one of the expendable two-factor interactions, say BC, with the blend differences. The table of coefficients is as follows:

	A	B	C	B_1	B_2		
Treatment	**1**	**2**	**3**	**4**	**5**	Group	Blend
1	–	–	–	+	–	1	3
2	+	–	–	+	+	2	4
3	–	+	–	–	+	3	1
4	+	+	–	–	–	4	2
5	–	–	+	–	+	3	1
6	+	–	+	–	–	4	2
7	–	+	+	+	–	1	3
8	+	+	+	+	+	2	4

The columns labelled B_1 and B_2 are just the columns of ± for BC and ABC; that is, the arrangement is generated by B_1 = BC and B_2 = ABC. The treatments in a particular group are determined by placing those with the same combination of ± for B_1 and B_2 in the same group. The 4 groups are then randomized to the 4 blends.

There is a serious weakness with this design!!! There are 3 degrees of freedom associated with group differences and we know of only two degrees of freedom confounded with Blends. What has happened to the third degree of freedom? Well, it is obtained as the interaction of B_1 and B_2. It will be found if you multiply these columns together you obtain the A column. Disaster! a main effect has been confounded with Blends. The experimental structure and analysis of variance table for this experiment are:

Structure	Formula
unrandomized	*4* Blends/*2* Runs
randomized	*2* A**2* B**2* C

Source	df	E[MSq]
Blends	3	
A	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{A}(\psi)$
B.C	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{B.C}(\psi)$
A.B.C	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{A.B.C}(\psi)$
Blends.Runs	4	
B	1	$\sigma^2_{BR} + f_{B}(\psi)$
A.B	1	$\sigma^2_{BR} + f_{A.B}(\psi)$
C	1	$\sigma^2_{BR} + f_{C}(\psi)$
A.C	1	$\sigma^2_{BR} + f_{A.C}(\psi)$

Fortunately, a calculus is available for avoiding such traps.

Theorem X.2: Let the columns in a table of ±s whose rows specify the combinations of the factors in a two-factor experiment be numbered **1**, **2**, …, ***m***. Also, let **I** be the column consisting entirely of +s. The elementwise product of two columns is commutative, the elementwise product of a column with **I** is the column itself and the elementwise product of a column with itself is **I**; that is,

$$\mathbf{ij} = \mathbf{ji},\ \ \mathbf{Ii} = \mathbf{iI} = \mathbf{i} \text{ and } \mathbf{ii} = \mathbf{I} \text{ where } \mathbf{i},\mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{m}$$

Proof: follows directly from a consideration of the results of multiplying ±1s together■

Example X.5 Complete sets of factorial treatments in 4 blocks (continued)

Firstly, number the factors as shown in the table. Thus, we can write

$$\mathbf{I} = \mathbf{11} = \mathbf{22} = \mathbf{33} = \mathbf{44} = \mathbf{55}$$

Now in the blocking arrangement just considered **4** = **23** and **5** = **123**.

The **45** column is thus **45** = **23**.**123** = **12233** = **1II** = **1** which shows that **45** is identical to **1** and that the interaction **45** is confounded with **1**.

A better arrangement is obtained by confounding the two block variables with any two of the two-factor interactions. The third degree of freedom is then confounded with the third two-factor interaction. Thus, for **4** = **12**, **5** = **13**, the interaction **45** is confounded with **23** since **45** = **1123** = **23**.

The experimental arrangement is indicated in the following table:

Treatment	A **1**	B **2**	C **3**	B_1 **4**	B_2 **5**	Group
1	–	–	–	+	+	1
2	+	–	–	–	–	2
3	–	+	–	–	+	3
4	+	+	–	+	–	4
5	–	–	+	+	–	4
6	+	–	+	–	+	3
7	–	+	+	–	–	2
8	+	+	+	+	+	1

The groups would be randomized to the blends and the order of the two runs for each blend would be randomized for each blend.

The analysis of variance table for the experiment (same structure as before) is:

Source	df	E[MSq]
Blends	3	
A.B	1	$\sigma^2_{BR} + 2\sigma^2_B + f_{A.B}(\psi)$
A.C	1	$\sigma^2_{BR} + 2\sigma^2_B + f_{A.C}(\psi)$
B.C	1	$\sigma^2_{BR} + 2\sigma^2_B + f_{B.C}(\psi)$
Blends.Runs	4	
A	1	$\sigma^2_{BR} + f_{A}(\psi)$
B	1	$\sigma^2_{BR} + f_{B}(\psi)$
C	1	$\sigma^2_{BR} + f_{C}(\psi)$
A.B.C	1	$\sigma^2_{BR} + f_{A.B.C}(\psi)$

The two runs for each group are complimentary in the sense that the plus minus pattern of one run is precisely the opposite to that of the second run. For example, in group 2, the plus and minus signs for the pairs of runs are +–– and –++.

Definition X.11: Two factor combinations are called a **fold-over pair** if the signs for the factors in one combination are exactly the opposite of those in the other combination. ■

Any 2^k factorial may be broken into 2^{k-1} blocks of size 2 by forming blocks such that each of them consists of a different fold-over pair. Such blocking arrangements leave the main effects of the *k* factors unconfounded with blocks. However, all two factor interactions are confounded with blocks. A layout for such designs cannot be produced with Genstat's *Stats > Design > Select Design* command.

Example X.6: Repeated four block experiment

As before we could replicate the whole experiment so that there are 4*r* blends. The 4 groups of treatments might then be assigned completely at random or in blocks. Suppose that the groups have been assigned completely at random to the 4*r* blends so that each group occurs *r* times. The analysis would then be:

Source	df	E[MSq]
Blends	$4r-1$	
A.B	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{A.B}(\psi)$
A.C	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{A.C}(\psi)$
B.C	1	$\sigma^2_{BR} + 2\sigma^2_{B} + f_{B.C}(\psi)$
Residual	$4(r-1)$	$\sigma^2_{BR} + 2\sigma^2_{B}$
Blends.Runs	$4r$	
A	1	$\sigma^2_{BR} + f_{A}(\psi)$
B	1	$\sigma^2_{BR} + f_{B}(\psi)$
C	1	$\sigma^2_{BR} + f_{C}(\psi)$
A.B.C	1	$\sigma^2_{BR} + f_{A.B.C}(\psi)$
Residual	$4(r-1)$	σ^2_{BR}

The expected mean squares for this experiment are based on equating the unrandomized and randomized factors with random and fixed factors, respectively. There are thus two sources of uncontrolled variation: differences between blends

$\left(\sigma^2_{\mathrm{BR}}\right)$ and between runs within blends $\left(\sigma^2_{\mathrm{B}}\right)$. The analysis indicates that the two-factor interactions are going to be affected by blend differences whereas the other effects will not. As blends are likely to be more variable than runs within blends, if blocking is effective as planned, the two-factor interactions are determined with less precision than the other effects. This will be a problem if it is anticipated that there are likely to be two-factor interactions; in such circumstances one needs to consider partial confounding. Again, a layout for such designs cannot be produced with Genstat's *Stats > Design > Select Design* command.

b) Partial confounding of effects

In experiments where the complete set of treatments are replicated it is possible to confound different effects in each replicate. This is called *partial confounding*.

Definition X.12: **Partial confounding** occurs when the effects confounded between blocks is different for different groups of blocks

Example X.7 Partial confounding in a repeated four block experiment

Suppose that we are wanting to run a four block experiment with repeats such as that discussed in example X.6. The total confounding options discussed for that example may be unsatisfactory because of the need to confound two-factor interactions. The use of partial confounding is investigated for this experiment. Consider the following generators for an experiment involving sets of 4 blocks:

Set	Group generators		
I	**4 = 12**	**5 = 13**	**45 = 23**
II	**4 = 123**	**5 = 23**	**45 = 1**
III	**4 = 123**	**5 = 13**	**45 = 2**
IV	**4 = 123**	**5 = 12**	**45 = 3**

Thus the three factor interaction is confounded in three sets, the two factor interactions in 2 sets and the main effects in 1 set. The formation of the groups of treatments is shown in the following table.

Set	Treatment	A **1**	B **2**	C **3**	B_1 **4**	B_2 **5**	Group
					12	**13**	
	1	–	–	–	+	+	1
	2	+	–	–	–	–	2
	3	–	+	–	–	+	3
	4	+	+	–	+	–	4
I	5	–	–	+	+	–	4
	6	+	–	+	–	+	3
	7	–	+	+	–	–	2
	8	+	+	+	+	+	1
					23	**123**	
	1	–	–	–	+	–	5
	2	+	–	–	+	+	6
	3	–	+	–	–	+	7
	4	+	+	–	–	–	8
II	5	–	–	+	–	+	7
	6	+	–	+	–	–	8
	7	–	+	+	+	–	5
	8	+	+	+	+	+	6
					13	**123**	
	1	–	–	–	+	–	9
	2	+	–	–	–	+	10
	3	–	+	–	+	+	11
	4	+	+	–	–	–	12
III	5	–	–	+	–	+	10
	6	+	–	+	+	–	9
	7	–	+	+	–	–	12
	8	+	+	+	+	+	11
					12	**123**	
	1	–	–	–	+	–	13
	2	+	–	–	–	+	14
	3	–	+	–	–	+	14
	4	+	+	–	+	–	13
IV	5	–	–	+	+	+	15
	6	+	–	+	–	–	16
	7	–	+	+	–	–	16
	8	+	+	+	+	+	15

The groups (pairs) of treatments in this table would then be randomized to the blends and the two treatment combinations in each group randomized to the two runs made for each blend. The layout and data for the experiment are shown in the table below. A layout for such a design cannot be produced with Genstat's *Stats > Design > Select Design* command.

Blends	Runs	Group	A	B	C	Yield
1	1	15	–	–	+	54.40
	2		+	+	+	81.40
2	1	10	+	–	–	70.60
	2		–	–	+	46.60
3	1	5	–	–	–	61.50
	2		–	+	+	44.50
4	1	9	–	–	–	60.70
	2		+	–	+	82.70
5	1	2	–	+	+	51.90
	2		+	–	–	79.90
6	1	16	–	+	+	43.90
	2		+	–	+	84.90
7	1	7	–	–	+	55.80
	2		–	+	–	59.80
8	1	14	+	–	–	73.50
	2		–	+	–	61.50
9	1	3	+	–	+	81.50
	2		–	+	–	50.50
10	1	11	+	+	+	80.90
	2		–	+	–	51.90
11	1	12	+	+	–	76.40
	2		–	+	+	53.40
12	1	8	+	+	–	68.60
	2		+	–	+	86.60
13	1	13	–	–	–	63.90
	2		+	+	–	69.90
14	1	6	+	–	–	75.90
	2		+	+	+	86.90
15	1	1	+	+	+	83.30
	2		–	–	–	63.30
16	1	4	+	+	–	63.50
	2		–	–	+	44.50

The experimental structure for this experiment is:

Structure	Formula
unrandomized	*16* Blends/*2* Runs
randomized	*2* A**2* B**2* C

The analysis of variance table for this experiment is:

Source	df	MSq	E[MSq]	F	Prob
Blends	15				
A	1	1161.62	$\sigma_{BR}^2 + 2\sigma_B^2 + f_A(\psi)$	34.55	<.001
B	1	0.50	$\sigma_{BR}^2 + 2\sigma_B^2 + f_B(\psi)$	0.01	0.906
C	1	2.20	$\sigma_{BR}^2 + 2\sigma_B^2 + f_C(\psi)$	0.07	0.804
A.B	1	0.72	$\sigma_{BR}^2 + 2\sigma_B^2 + f_{A.B}(\psi)$	0.02	0.887
A.C	1	289.00	$\sigma_{BR}^2 + 2\sigma_B^2 + f_{A.C}(\psi)$	8.60	0.019
B.C	1	82.81	$\sigma_{BR}^2 + 2\sigma_B^2 + f_{B.C}(\psi)$	2.46	0.155
A.B.C	1	0.20	$\sigma_{BR}^2 + 2\sigma_B^2 + f_{A.B.C}(\psi)$	0.01	0.940
Residual	8	33.62	$\sigma_{BR}^2 + 2\sigma_B^2$		
Blends.Runs	16				
A	1	3313.50	$\sigma_{BR}^2 + f_A(\psi)$	1016.64	<0.001
B	1	150.00	$\sigma_{BR}^2 + f_B(\psi)$	46.02	<0.001
C	1	10.67	$\sigma_{BR}^2 + f_C(\psi)$	3.27	0.104
A.B	1	9.00	$\sigma_{BR}^2 + f_{A.B}(\psi)$	2.76	0.131
A.C	1	625.00	$\sigma_{BR}^2 + f_{A.C}(\psi)$	191.76	<0.001
B.C	1	0.00	$\sigma_{BR}^2 + f_{B.C}(\psi)$	0.00	1.000
A.B.C	1	0.50	$\sigma_{BR}^2 + f_{A.B.C}(\psi)$	0.15	0.704
Residual	9	3.26	σ_{BR}^2		
Total	31				

Model term	e.f.	nonorthogonal terms
Information summary		
Blends stratum		
A	0.250	
B	0.250	
C	0.250	
A.B	0.500	
A.C	0.500	
B.C	0.500	
A.B.C	0.750	
Blends.Runs stratum		
A	0.750	Blends
B	0.750	Blends
C	0.750	Blends
A.B	0.500	Blends
A.C	0.500	Blends
B.C	0.500	Blends
A.B.C	0.250	Blends

Clearly, the experiment is balanced; the analysis can be accomplished using Yates algorithm, taking into account the efficiency with which various terms are estimated. Also, diagnostic checking would be performed on the residuals.

c) Computation in Genstat

In general, the analysis of confounded designs depends on whether only a single replicate of the treatments has been observed in the experiment or several replicates of the complete set of treatments have been observed.

If a single replicate of the treatments has been observed, then the analysis must be based on the normal probability plot of the Yates effects. On the other hand, if two or more complete sets of treatments have been observed, the analysis can be based on an analysis of variance table.

Example X.7 Partial confounding in a repeated four block experiment (continued)

As four complete sets of treatments have been observed, the analysis will use an analysis of variance table. Having set up the factors and data in a Genstat spreadsheet, the instructions to produce this analysis are as follows:

```
PRINT Blends,Runs,A,B,C,Yield
BLOCKS Blends/Runs
TREAT A*B*C
ANOVA [PRINT=A,I,E,M; TWOLEVEL=Yates; FPROB=Y; PSE=LSD] Yield
APLOT METHOD=fit,normal
"
**** Tukey''s one-degree-of-freedom-for-non-additivity.
**** It is the term designated covariate in the following analysis
"
AKEEP [FIT=Fit]
CALC ResSq=Fit*Fit
ANOVA [PRINT=*]  ResSq; RES=ResSq
COVAR ResSq                             "A computational trick"
ANOVA [PRINT=A; FPROB=Y] Yield
COVAR
```

X.D Fractional factorial designs at two levels

(Box, Hunter & Hunter, ch. 12)

The number of runs required by a full 2^k increases geometrically as k increases. However, it turns out that when k is not small the desired information can often be obtained by performing only a fraction of the full factorial design.

In previous sections it was suggested that there was a great deal of redundancy in a factorial experiment in that higher-order interactions are likely to be negligible and some variables may not affect the response at all. We utilized this fact to suggest that it was not necessary to replicate the various treatments. In this section we go one step further by saying that you need take only a fraction of the full factorial design.

Consider a 2^7 design. A complete factorial arrangement requires $2^7 = 128$ runs. From these runs 128 effects can be calculated as follows:

				interactions of order			
average	main effects	2	3	4	5	6	7
1	7	21	35	35	21	7	1

So there tends to be redundancy in 2^k designs in that there is likely to be an excess number of interactions that can be estimated and sometimes an excess number of variables studied. Fractional factorial designs exploit this redundancy. To illustrate the ideas the example given by Box, Hunter and Hunter will be presented. It involves a half-fraction of a 2^5.

Example X.8 A complete 2^5 factorial experiment

The data from a complete 2^5 is given in the table below. Note that the results are not in the randomized order that would be order in which the runs were actually conducted. To make it easier to see how the design was constructed, the order of the runs is such that the treatments are in standard order.

Results from a 2^5 factorial design — chemical experiment

Factor	-	+
1 feed rate (l/min)	10	15
2 catalyst (%)	1	2
3 agitation rate (rpm)	100	120
4 temperature (°C)	140	180
5 concentration (%)	3	6

			Factor			
Run	**1**	**2**	**3**	**4**	**5**	% reacted
1	–	–	–	–	–	61
*2	+	–	–	–	–	53
*3	–	+	–	–	–	63
4	+	+	–	–	–	61
*5	–	–	+	–	–	53
6	+	–	+	–	–	56
7	–	+	+	–	–	54
*8	+	+	+	–	–	61
*9	–	–	–	+		69
10	+	–	–	+	–	61
11	–	+	–	+	–	94
*12	+	+	–	+	–	93
13	–	–	+	+	–	66
*14	+	–	+	+	–	60
*15	–	+	+	+	–	95
16	+	+	+	+	–	98
*17	–	–	–	–	+	56
18	+	–	–	–	+	63
19	–	+	–	–	+	70
*20	+	+	–	–	+	65
21	–	–	+	–	+	59
*22	+	–	+	–	+	55
*23	–	+	+	–	+	67
24	+	+	+	–	+	65
25	–	–	–	+	+	44
*26	+	–	–	+	+	45
*27	–	+	–	+	+	78
28	+	+	–	+	+	77
*29	–	–	+	+	+	49
30	+	–	+	+	+	42
31	–	+	+	+	+	81
*32	+	+	+	+	+	82

The experimental structure for this experiment is the standard structure for a 2^5 CRD; it is:

Structure	Formula
unrandomized	*32* Runs
randomized	*2* Feed**2* Catal**2* Agitation**2* Temp**2* Conc

The Genstat output file for the analysis of this example is as follows:

```
Genstat 5  Release 4.1  (PC/Windows NT)             10 April 2000 17:58:40
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                  ______________________________________________

                  Genstat 5 Fourth Edition - (for Windows)
                  Genstat 5 Procedure Library Release PL11
                  ______________________________________________

   3  "Data taken from File: D:/ANALYSES/LM/MULTIFAC/FAC5REAC.GSH"
   4  DELETE [redefine=yes] Runs,Feed,Catal,Agitatio,Temp,Conc,%Reacted
   5  FACTOR [modify=yes;nvalues=32;levels=32] Runs
   6  READ Runs; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Runs        32         0        32

   9  FACTOR [modify=yes;nvalues=32;levels=2] Feed
  10  READ Feed; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Feed        32         0         2

  12  FACTOR [modify=yes;nvalues=32;levels=2] Catal
  13  READ Catal; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Catal        32         0         2

  15  FACTOR [modify=yes;nvalues=32;levels=2] Agitatio
  16  READ Agitatio; frepresentation=ordinal

    Identifier    Values   Missing    Levels
      Agitatio        32         0         2

  18  FACTOR [modify=yes;nvalues=32;levels=2] Temp
  19  READ Temp; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Temp        32         0         2

  21  FACTOR [modify=yes;nvalues=32;levels=2] Conc
  22  READ Conc; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Conc        32         0         2

  24  VARIATE [nvalues=32] %Reacted
  25  READ %Reacted

    Identifier   Minimum      Mean   Maximum    Values   Missing
      %Reacted     42.00     65.50     98.00        32         0

  28
  29  PRINT Runs,Feed,Catal,Agitation,Temp,Conc,%Reacted; FIELD=10; DEC=0

      Runs      Feed     Catal  Agitatio      Temp      Conc  %Reacted
         1         1         1         1         1         1        61
         2         2         1         1         1         1        53
```

3	1	2	1	1	1	63
4	2	2	1	1	1	61
5	1	1	2	1	1	53
6	2	1	2	1	1	56
7	1	2	2	1	1	54
8	2	2	2	1	1	61
9	1	1	1	2	1	69
10	2	1	1	2	1	61
11	1	2	1	2	1	94
12	2	2	1	2	1	93
13	1	1	2	2	1	66
14	2	1	2	2	1	60
15	1	2	2	2	1	95
16	2	2	2	2	1	98
17	1	1	1	1	2	56
18	2	1	1	1	2	63
19	1	2	1	1	2	70
20	2	2	1	1	2	65
21	1	1	2	1	2	59
22	2	1	2	1	2	55
23	1	2	2	1	2	67
24	2	2	2	1	2	65
25	1	1	1	2	2	44
26	2	1	1	2	2	45
27	1	2	1	2	2	78
28	2	2	1	2	2	77
29	1	1	2	2	2	49
30	2	1	2	2	2	42
31	1	2	2	2	2	81
32	2	2	2	2	2	82

```
BLOCK Runs
TREAT Feed*Catal*Agitation*Temp*Conc
"Produce normal plot of Yates effects"
A2PLOT [PRINT=E; FACTORIAL=4; STRATUM=Runs; METHOD=normal; \
        GRAPH=line] %Reacted
```

```
34...............................................................................

***** Tables of effects *****

Variate: %Reacted

***** Runs stratum *****

Feed Y-effect               -1.37   s.e. 0.500     rep. 16

Catal Y-effect              19.50   s.e. 0.500     rep. 16

Agitatio Y-effect           -0.63   s.e. 0.500     rep. 16

Temp Y-effect               10.75   s.e. 0.500     rep. 16

Conc Y-effect               -6.25   s.e. 0.500     rep. 16

Feed.Catal Y-effect          1.37   s.e. 0.500     rep. 8

Feed.Agitatio Y-effect       0.75   s.e. 0.500     rep. 8

Catal.Agitatio Y-effect      0.88   s.e. 0.500     rep. 8

Feed.Temp Y-effect          -0.88   s.e. 0.500     rep. 8

Catal.Temp Y-effect         13.25   s.e. 0.500     rep. 8

Agitatio.Temp Y-effect       2.12   s.e. 0.500     rep. 8

Feed.Conc Y-effect           0.13   s.e. 0.500     rep. 8

Catal.Conc Y-effect          2.00   s.e. 0.500     rep. 8
```

```
Agitatio.Conc Y-effect      0.88   s.e. 0.500     rep. 8

Temp.Conc Y-effect        -11.00   s.e. 0.500     rep. 8

Feed.Catal.Agitatio Y-effect
                            1.50   s.e. 0.500     rep. 4

Feed.Catal.Temp Y-effect
                            1.37   s.e. 0.500     rep. 4

Feed.Agitatio.Temp Y-effect
                           -0.75   s.e. 0.500     rep. 4

Catal.Agitatio.Temp Y-effect
                            1.12   s.e. 0.500     rep. 4

Feed.Catal.Conc Y-effect
                           -1.87   s.e. 0.500     rep. 4

Feed.Agitatio.Conc Y-effect
                           -2.50   s.e. 0.500     rep. 4

Catal.Agitatio.Conc Y-effect
                            0.13   s.e. 0.500     rep. 4

Feed.Temp.Conc Y-effect     0.63   s.e. 0.500     rep. 4

Catal.Temp.Conc Y-effect
                           -0.25   s.e. 0.500     rep. 4

Agitatio.Temp.Conc Y-effect
                            0.13   s.e. 0.500     rep. 4

Feed.Catal.Agitatio.Temp Y-effect
                            0.00   s.e. 0.500     rep. 2

Feed.Catal.Agitatio.Conc Y-effect
                            1.50   s.e. 0.500     rep. 2

Feed.Catal.Temp.Conc Y-effect
                            0.63   s.e. 0.500     rep. 2

Feed.Agitatio.Temp.Conc Y-effect
                            1.00   s.e. 0.500     rep. 2

Catal.Agitatio.Temp.Conc Y-effect
                           -0.63   s.e. 0.500     rep. 2
```

```
             -
N            I
o            I                                                          *
r     1.6031 I                                             *
m            I                         *               *
a            I                        **
l            I                        3
             I                       3*
Q            I                      *3
u    -0.3206 I                     *3
a            I                    *2
n            I                   **
t            I                 **
i            I          *
l            I  *
e    -2.2443 I
             -+---------+---------+---------+---------+---------+---------+-
        -12.100    -6.508    -0.917     4.675    10.267    15.858    21.450

                                            effect
```

```
"Perform analysis including only significant effects
   and do Residual analysis"
BLOCK Runs
TREAT (Catal+Conc)*Temp
ANOVA [FPROB=Y; PSE=LSD] %Reacted; RES=Res
```

40...

***** Analysis of variance *****

Variate: %Reacted

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Runs stratum					
Catal	1	3042.00	3042.00	274.15	<.001
Conc	1	312.50	312.50	28.16	<.001
Temp	1	924.50	924.50	83.32	<.001
Catal.Temp	1	1404.50	1404.50	126.58	<.001
Conc.Temp	1	968.00	968.00	87.24	<.001
Residual	26	288.50	11.10		
Total	31	6940.00			

***** Tables of means *****

Variate: %Reacted

Grand mean 65.50

Catal	1	2
	55.75	75.25

Conc	1	2
	68.62	62.38

Temp	1	2
	60.13	70.87

Catal	Temp	1	2
1		57.00	54.50
2		63.25	87.25

Conc	Temp	1	2
1		57.75	79.50
2		62.50	62.25

*** Least significant differences of means (5% level) ***

Table	Catal	Conc	Temp	Catal Temp
rep.	16	16	16	8
d.f.	26	26	26	26
l.s.d.	2.421	2.421	2.421	3.424

Table	Conc Temp
rep.	8
d.f.	26
l.s.d.	3.424

```
  41  APLOT METHOD=fit,normal


             -
             I
             I
             I
         8.0 I
r            I               *
e            I                       *
s            I                   *      *            *
i            I      *              * *               *                *
d            I               *          *
u        0.0 I                   * 2    2            *
a            I      2        2                       *                2
l            I                       2                                *
s            I      *            2
             I
             I                     *
        -8.0 I
             -+---------+---------+---------+---------+---------+---------+---
           40.0      50.0      60.0      70.0      80.0      90.0     100.0

                                         fitted values


                                          Normal plot

             -
             I
         8.0 I
r            I                                                        *
e            I                                                   *
s            I                                           * * *
i            I                                     **** *
d            I                                  * *
u        0.0 I                             *2***
a            I                     **** ***
l            I                 ** *
s            I         *   * *
             I
             I    *
        -8.0 I
             -+---------+---------+---------+---------+---------+---------+---
           -2.4      -1.6      -0.8       0.0       0.8       1.6       2.4

                                   expected Normal quantiles

  42  "
 -43  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -44  **** It is the term designated covariate in the following analysis
 -45  "
  46  AKEEP [FIT=Fit]
  47  CALC ResSq=Fit*Fit
  48  ANOVA [PRINT=*]  ResSq; RES=ResSq
  49  COVAR ResSq                            "A computational trick"
  50  ANOVA [PRINT=A; FPROB=Y] %Reacted
```

```
50.......................................................................

***** Analysis of variance (adjusted for covariate) *****

Variate: %Reacted
Covariate: ResSq

Source of variation     d.f.       s.s.       m.s.    v.r. cov.ef.  F pr.

Runs stratum
Catal                      1    3042.00    3042.00  273.68    1.00  <.001
Conc                       1     312.50     312.50   28.11    1.00  <.001
Temp                       1     924.50     924.50   83.17    1.00  <.001
Catal.Temp                 1    1404.50    1404.50  126.36    1.00  <.001
Conc.Temp                  1     968.00     968.00   87.09    1.00  <.001
Covariate                  1      10.62      10.62    0.96          0.338
Residual                  25     277.88      11.12            1.00
Total                     31    6940.00

  51  COVAR
```

From this we conclude that the main effects Catal, Temp and Conc and the two-factor interactions Catal.Temp and Temp.Conc are the only effects distinguishable from noise. The plot of the residuals is fine and so also is the normal probability plot. Tukey's one-degree-of-freedom-for-nonadditivity is not significant. So there is no evidence that the assumptions are unmet.

a) Half-fractions of full factorial experiments

Definition X.13: A $\frac{1}{p}$th fraction of a 2^k experiment is designated a $\mathbf{2^{k\text{-}p}}$ **experiment**. The number of runs in the experiment is equal to the value of $2^{k\text{-}p}$. ■

Construction of half-fractions

Rule X.1: A $2^{k\text{-}1}$ experiment is constructed as follows:

1. Write down a complete design in k–1 factors.
2. Compute the column of signs for factor k by forming the elementwise product of the columns of the complete design. That is, $\mathbf{k} = \mathbf{123}\ldots(\mathbf{k}-\mathbf{1})$. ■

Example X.9 A half-fraction of a 2^5 factorial experiment

Now, the full factorial experiment analysed in example X.8 required 32 runs. Suppose that the experimenter had chosen to make only the 16 runs marked with asterisks in the above table — that is, the $2^4 = 16$ runs specified by rule X.1 for a $2^{5\text{-}1}$ design:

1. A full 2^4 design was chosen for the four factors **1**, **2**, **3** and **4**.
2. The column of signs for the four-factor interaction was computed and these were used to define the levels of factor **5**. Thus, $\mathbf{5} = \mathbf{1234}$.

Consequently, the only data available would be that given in the table below; this data has been arranged in standard order for factors 1–4, not in randomized order. Also, given are the coefficients of the contrasts for all the two-factor interactions.

Results from a half-fraction of a 2^5 factorial design — chemical experiment

Factor		–	+
1	feed rate (l/min)	10	15
2	catalyst (%)	1	2
3	agitation rate (rpm)	100	120
4	temperature (°C)	140	180
5	concentration (%)	3	6

	Factor					Interactions										%
Run	**1**	**2**	**3**	**4**	**5**	**12**	**13**	**14**	**15**	**23**	**24**	**25**	**34**	**35**	**45**	reacted
17	–	–	–	–	+	+	+	+	–	+	+	–	+	–	–	56
2	+	–	–	–	–	–	–	–	–	+	+	+	+	+	+	53
3	–	+	–	–	–	–	+	+	+	–	–	–	+	+	+	63
20	+	+	–	–	+	+	–	–	+	–	–	+	+	–	–	65
5	–	–	+	–	–	+	–	+	+	–	+	+	–	–	+	53
22	+	–	+	–	+	–	+	–	+	–	+	–	–	+	–	55
23	–	+	+	–	+	–	–	+	–	+	–	+	–	+	–	67
8	+	+	+	–	–	+	+	–	–	+	–	–	–	–	+	61
9	–	–	–	+	–	+	+	–	+	+	–	+	–	+	–	69
26	+	–	–	+	+	+	–	+	+	+	–	–	–	–	+	45
27	–	+	–	+	+	–	+	–	–	–	+	+	–	–	+	78
12	+	+	–	+	–	+	–	+	–	–	+	–	–	+	–	93
29	–	–	+	+	+	+	–	–	–	–	–	–	+	+	+	49
14	+	–	+	+	–	–	+	+	–	–	–	+	+	–	–	60
15	–	+	+	+	–	–	–	–	+	+	+	–	+	–	–	95
32	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	82

Aliasing in half-fractions

As one would expect there is no such thing as a free lunch. We have gained by only having to run half the combinations, but what has been lost? The short answer is that various effects have been aliased.

Definition X.14: Two effects are said to be **aliased** when they are mixed up because of the deliberate use of only a fraction of the treatments. ■

Compare this to confounding, where certain treatment effects are mixed up with block effects. That is, the inability to separate the effects arises from different actions. In one case, it is because of the treatment combinations that the investigator chooses to observe; in the other case, it arises from the assigning of treatments to physical units.

In the table above, only the columns for the main effects and two-factor interactions are presented. What about the 10 three-factor interactions, the 5 four-factor interactions and the 1 five-factor interaction? Consider the three factor interaction **123**; its coefficients are:

123 = –++–+––+–++–+––+

but we notice this is identical to the column **45** in the above table. That is, **123** = **45** and these two interactions are aliased.

Now suppose we use ℓ_{45} to denote the linear function of the observations which we used to estimate the **45** interaction:

$$\ell_{45} = (-56+53+63-65+53-55-67+61-69+45+78-93+49-60-95+82)/8 = -9.5$$

Now, ℓ_{45} estimates the sum of the effects **45** and **123** from the complete design. It is said that ℓ_{45} ➔ **45** + **123**. That is, it is the sum of the parameters for **45** and **123** that is estimated by ℓ_{45}.The complete aliasing pattern for this design is as follows:

Aliasing pattern for a 2^{5-1} design

Relationship between column pairs			Aliasing pattern				
1	=	**2345**	ℓ_1	➔	**1**	+	**2345**
2	=	**1345**	ℓ_2	➔	**2**	+	**1345**
3	=	**1245**	ℓ_3	➔	**3**	+	**1245**
4	=	**1235**	ℓ_4	➔	**4**	+	**1235**
5	=	**1234**	ℓ_5	➔	**5**	+	**1234**
12	=	**345**	ℓ_{12}	➔	**12**	+	**345**
13	=	**245**	ℓ_{13}	➔	**13**	+	**245**
14	=	**235**	ℓ_{14}	➔	**14**	+	**235**
15	=	**234**	ℓ_{15}	➔	**15**	+	**234**
23	=	**145**	ℓ_{23}	➔	**23**	+	**145**
24	=	**135**	ℓ_{24}	➔	**24**	+	**135**
25	=	1**34**	ℓ_{25}	➔	**25**	+	**134**
34	=	**125**	ℓ_{34}	➔	**34**	+	**125**
35	=	**124**	ℓ_{35}	➔	**35**	+	**124**
45	=	**123**	ℓ_{45}	➔	**45**	+	**123**
(**I**	=	**12345**)	[ℓ_I	➔	average + ½(**12345**)]		

Evidently our analysis would be justified if it could be assumed that the three-factor and four-factor interactions could be ignored.

Analysis of half-fractions

The analysis of this set of 16 runs can still be accomplished using Yates algorithm since there are 4 factors for which it represents a full factorial. However, Genstat will perform the analysis producing lines for a set of unaliased terms. The experimental structure is the same as for the full factorial. The Genstat output file for the analysis is as follows:

```
Genstat 5  Release 4.1  (PC/Windows NT)              10 April 2000 21:09:03
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

                 Genstat 5 Fourth Edition - (for Windows)
                 Genstat 5 Procedure Library Release PL11
                 ______________________________________________

   3  "Data taken from File: D:/ANALYSES/LM/MULTIFAC/FRF5REAC.GSH"
   4  DELETE [redefine=yes] Runs,Feed,Catal,Agitatio,Temp,Conc,%Reacted
   5  FACTOR [modify=yes;nvalues=16;levels=16] Runs
   6  READ Runs; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Runs        16         0        16

   8  FACTOR [modify=yes;nvalues=16;levels=2] Feed
   9  READ Feed; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Feed        16         0         2

  11  FACTOR [modify=yes;nvalues=16;levels=2] Catal
  12  READ Catal; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Catal        16         0         2

  14  FACTOR [modify=yes;nvalues=16;levels=2] Agitatio
  15  READ Agitatio; frepresentation=ordinal

    Identifier    Values   Missing    Levels
      Agitatio        16         0         2

  17  FACTOR [modify=yes;nvalues=16;levels=2] Temp
  18  READ Temp; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Temp        16         0         2

  20  FACTOR [modify=yes;nvalues=16;levels=2] Conc
  21  READ Conc; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Conc        16         0         2

  23  VARIATE [nvalues=16] %Reacted
  24  READ %Reacted

    Identifier   Minimum      Mean   Maximum    Values   Missing
      %Reacted     45.00     65.25     95.00        16         0

  26
```

```
  27  PRINT Runs,Feed,Catal,Agitation,Temp,Conc,%Reacted; FIELD=10; DEC=0

      Runs      Feed     Catal  Agitatio      Temp      Conc  %Reacted
         1         1         1         1         1         2        56
         2         2         1         1         1         1        53
         3         1         2         1         1         1        63
         4         2         2         1         1         2        65
         5         1         1         2         1         1        53
         6         2         1         2         1         2        55
         7         1         2         2         1         2        67
         8         2         2         2         1         1        61
         9         1         1         1         2         1        69
        10         2         1         1         2         2        45
        11         1         2         1         2         2        78
        12         2         2         1         2         1        93
        13         1         1         2         2         2        49
        14         2         1         2         2         1        60
        15         1         2         2         2         1        95
        16         2         2         2         2         2        82

  28  BLOCK Runs
  29  TREAT Feed*Catal*Agitation*Temp*Conc
  30  "Produce normal plot of Yates effects"
  31  A2PLOT [PRINT=inform,effect; FACTORIAL=4; STRATUM=Runs; METHOD=normal; \
  32          GRAPH=line] %Reacted


32...............................................................................

***** Information summary *****

Aliased model terms
Feed.Catal.Agitatio
Feed.Catal.Temp
Feed.Agitatio.Temp
Catal.Agitatio.Temp
Feed.Catal.Conc
Feed.Agitatio.Conc
Catal.Agitatio.Conc
Feed.Temp.Conc
Catal.Temp.Conc
Agitatio.Temp.Conc
Feed.Catal.Agitatio.Temp
Feed.Catal.Agitatio.Conc
Feed.Catal.Temp.Conc
Feed.Agitatio.Temp.Conc
Catal.Agitatio.Temp.Conc


***** Tables of effects *****

Variate: %Reacted

***** Runs stratum *****

Feed Y-effect              -2.00   s.e. *         rep. 8

Catal Y-effect             20.50   s.e. *         rep. 8

Agitatio Y-effect           0.00   s.e. *         rep. 8

Temp Y-effect              12.25   s.e. *         rep. 8

Conc Y-effect              -6.25   s.e. *         rep. 8

Feed.Catal Y-effect         1.50   s.e. *         rep. 4

Feed.Agitatio Y-effect      0.50   s.e. *         rep. 4

Catal.Agitatio Y-effect     1.50   s.e. *         rep. 4

Feed.Temp Y-effect         -0.75   s.e. *         rep. 4
```

```
Catal.Temp Y-effect        10.75   s.e. *         rep. 4

Agitatio.Temp Y-effect      0.25   s.e. *         rep. 4

Feed.Conc Y-effect          1.25   s.e. *         rep. 4

Catal.Conc Y-effect         1.25   s.e. *         rep. 4

Agitatio.Conc Y-effect      2.25   s.e. *         rep. 4

Temp.Conc Y-effect         -9.50   s.e. *         rep. 4

             -
N            I
o            I                                                        *
r     1.3667 I                                         *
m            I
a            I                       *               *
l            I                      *
             I                     **
Q            I                     *
u    -0.2733 I                   **
a            I                   *
n            I               *  *
t            I
i            I        *
l            I  *
e    -1.9133 I
             -+---------+---------+---------+---------+---------+---------+-
         -10.45     -4.95      0.55      6.05     11.55     17.05     22.55

                                            effect
```

The values of the Yates effects for the fractional set are similar to those obtained in the analysis of the complete set. Furthermore, the normal plot draws attention to precisely the same set of effects: Conc, Temp, Catal, Temp.Conc and Temp.Catal. Thus essentially the same information has been obtained with only half the effort.

The following Genstat output file gives the analysis for the significant effects (and effects marginal to them):

```
  33
  34  "Perform analysis including only significant effects
 -35     and do Residual analysis"
  36  BLOCK Runs
  37  TREAT (Catal+Conc)*Temp
  38  ANOVA [FPROB=Y; PSE=LSD] %Reacted; RES=Res


38..............................................................................

***** Analysis of variance *****

Variate: %Reacted

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Runs stratum
Catal                     1   1681.000   1681.000  239.29  <.001
Conc                      1    156.250    156.250   22.24  <.001
Temp                      1    600.250    600.250   85.44  <.001
Catal.Temp                1    462.250    462.250   65.80  <.001
Conc.Temp                 1    361.000    361.000   51.39  <.001
Residual                 10     70.250      7.025
Total                    15   3331.000
```

```
***** Tables of means *****

Variate: %Reacted

Grand mean  65.25

    Catal        1        2
             55.00    75.50

     Conc        1        2
             68.37    62.13

     Temp        1        2
             59.13    71.37

    Catal     Temp        1        2
        1             54.25    55.75
        2             64.00    87.00

     Conc     Temp        1        2
        1             57.50    79.25
        2             60.75    63.50


*** Least significant differences of means (5% level) ***

Table                 Catal        Conc        Temp       Catal
                                                           Temp
rep.                      8           8           8           4
d.f.                     10          10          10          10
l.s.d.                2.953       2.953       2.953       4.176

Table                  Conc
                       Temp
rep.                      4
d.f.                     10
l.s.d.                4.176


  39  APLOT METHOD=fit,normal


             -
             I
             I                        *
             I
         4.0 I
r            I
e            I                                       *
s            I
i            I        *                 *
d            I             2        *
u        0.0 I                *                                      *
a            I                *         *
l            I                      *                *
s            I                                                       *
             I        *
             I                        *
        -4.0 I
             -+---------+---------+---------+---------+---------+---------+---
           40.0      50.0      60.0      70.0      80.0      90.0     100.0

                                         fitted values


                                          Normal plot
```

```
             -
             I
         5.0 I                                                           *
r            I
e            I
s            I                                                   *
i            I                                              *
d            I                                       *   *
u        0.0 I                             * *  *  *
a            I                     * *  *
l            I              *  *
s            I         *
             I *
             I
        -5.0 I
             -+---------+---------+---------+---------+---------+---------+---
           -1.8      -1.2      -0.6       0.0       0.6       1.2       1.8

                                   expected Normal quantiles

  40  "
 -41  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -42  **** It is the term designated covariate in the following analysis
 -43  "
  44  AKEEP [FIT=Fit]
  45  CALC ResSq=Fit*Fit
  46  ANOVA [PRINT=*]  ResSq; RES=ResSq
  47  COVAR ResSq                        "A computational trick"
  48  ANOVA [PRINT=A; FPROB=Y] %Reacted


48...............................................................................

***** Analysis of variance (adjusted for covariate) *****

Variate: %Reacted
Covariate: ResSq

Source of variation     d.f.       s.s.       m.s.    v.r. cov.ef.  F pr.

Runs stratum
Catal                     1   1681.000   1681.000  234.62    1.00  <.001
Conc                      1    156.250    156.250   21.81    1.00  0.001
Temp                      1    600.250    600.250   83.78    1.00  <.001
Catal.Temp                1    462.250    462.250   64.52    1.00  <.001
Conc.Temp                 1    361.000    361.000   50.39    1.00  <.001
Covariate                 1      5.768      5.768    0.81          0.393
Residual                  9     64.482      7.165            0.98
Total                    15   3331.000

  49  COVAR
```

The diagnostic checking indicates the assumptions are satisfactory.

Finding the aliasing patterns

Fortunately, we can use the calculus, described in section X.C, for multiplying columns together to obtain the aliasing pattern.

Definition X.15: A set of columns designating an elementwise product of a set of columns is called a **word**. ■

Definition X.16: The **generator(s)** of a fractional 2^k experiment are the relationship(s) between factors that are used to obtain the design. The **generating relations** express these as words equal to **I**. ■

Definition X.17: The **defining relations** of a fractional 2^k experiment is the set of all words equal to the identity column **I**. It includes the generating relations and all possible products of these relations. ■

The defining relation(s) are the key to the design since multiplying them on both sides of the equals sign by particular column combinations yields the relationships between the columns.

Example X.9 A half-fraction of a 2^5 factorial experiment (continued)

The 2^{5-1} design was constructed by setting **5** = **1234**. That is, this relation is the generator of the design. Multiplying both sides by **5** we obtain **55** = **12345** or **I** = **12345**. This identity can be confirmed by multiplying together the columns in the table of coefficients for the design. The half fraction is defined by a single generating relation **I** = **12345** so that the relation **I** = **12345** also provides the defining relation of the design. (For more highly fractionated designs more than one defining relation is needed.)

To find out which column combination is aliased with **45**, multiply the defining relation on both sides by **45**: **45I** = **4512345** = **123** which is precisely the relation given in the above table. The other aliasing relations can be found similarly.

b) More on construction and use of half-factions

The complimentary half-fraction

Definition X.18: The **complimentary half-fraction** is obtained by reversing the sign of the generating relation. ■

In practice, either half-fraction may be used.

Example X.9 A half-fraction of a 2^5 factorial experiment (continued)

As already mentioned the design we have been discussing was constructed using the generator **5** = **1234**; that is, we formed the fifth column by multiplying together the other 4 columns. The complimentary half fraction is generated by putting **5** = –**1234**; that is, taking minus the product of the first 4 columns. The half fraction corresponding to the runs not marked with an asterisk in the complete experiment is obtained. The defining relation for this design may be written as **I** = –**12345**.

The complimentary half fraction would give:

$\ell_{\mathbf{1}} \Rightarrow \mathbf{1} - \mathbf{2345}$

$\ell_{\mathbf{2}} \Rightarrow \mathbf{2} - \mathbf{1345}$ etc.

c) The concept of design resolution

Definition X.19: The **resolution *R*** of a fractional design is the number of columns in the smallest word in the defining relations. ■

Thus a 2^{5-1} half fraction with defining relation **I** = ±**12345** has resolution V and is referred to as a 2_{V}^{5-1} design. A 2^{3-1} with defining relation **I** = ±**123** has resolution III and is referred to as a 2_{III}^{5-1} design.

Theorem X.3: The half-fraction of a 2^k experiment with highest resolution is the one for which the generator is $\mathbf{k} = \mathbf{123}\ldots(\mathbf{k}-\mathbf{1})$ and so the generating and defining relation is $\mathbf{I} = \mathbf{123}\ldots\mathbf{k}$.

Proof: it is clear that the largest word possible has been used for the generating relationship. ■

The half-fraction is the one that it was suggested as best in rule X.1.

Rule X.2: A design of resolution *R* is one in which *p*-factor effects may be aliased with effects containing *R* – *p* factors or more. ■

Example X.9 A half-fraction of a 2^5 factorial experiment (continued)

The 2^{5-1} design we have been discussing is a resolution V design and so all *p*-factor interactions are aliased with V–*p* interactions. Hence, all one-factor interactions are aliased with four-factor interactions and all two-factor interactions are aliased with three-factor interactions. ■

For the various sized fractions of a 2^k:

1. a design of resolution III does not alias main effects with one another, but does alias main effects with two factor interactions;

2. a design of resolution IV does not alias main effects and two-factor interactions, but does alias two factor interactions with one another;

3. a design of resolution V does not alias main effects and two-factor with one another, but does alias two factor interactions with three-factor interactions.

Embedded factorials

One of the very important aspects of resolution is that a fractional factorial design of resolution *R* contains complete factorials (possibly replicated) in every set of *R*–1 factors. Suppose then that an experimenter has a number of candidate factors but believes that all but *R*–1 of them (specific identity unknown) may have no detectable effects. If she employs a fractional factorial design of resolution *R* she will have a complete factorial design in the effective factors.

Example X.9 A half-fraction of a 2^5 factorial experiment (continued)

The 2_V^{5-1} design is a complete factorial in any four of the five factors. Thus, if we can identify a factor that is having no effect, the experiment can be interpreted as a complete factorial. In the analysis discussed above factors **1** (Feed) and **3** (Agitation) are without effect, so that the experiment can be viewed as a replicated 2^3 factorial experiment.

Results from a 2_V^{5-1} factorial design viewed as a 2^3 — chemical experiment

Factor	–	+
1 feed rate (l/min)	10	15
2 catalyst (%)	1	2
3 agitation rate (rpm)	100	120
4 temperature (°C)	140	180
5 concentration (%)	3	6

		Factor		
Run	**2**	**4**	**5**	% reacted
2	–	–	–	53
5	–	–	–	53
17	–	–	+	56
22	–	–	+	55
9	–	+	–	69
14	–	+	–	60
26	–	+	+	45
29	–	+	+	49
3	+	–	–	63
8	+	–	–	61
20	+	–	+	65
23	+	–	+	67
12	+	+	–	93
15	+	+	–	95
27	+	+	+	78
32	+	+	+	82

d) Resolution III designs (also called main effect plans)

In a resolution III design, the main effects are aliased with two-factor interactions. It is possible to construct resolution III designs for investigating up to $k = n–1$ factors, where n is a multiple of 4. A particularly important subset of these designs are those for which n is a power of 2. In particular, designs for investigating up to 3 factors in 4 runs, up to 7 factors in 8 runs and up to 15 factors in 16 runs.

Definition X.20: A fractional design is said to be a **saturated design** if the number of factors $k = n{-}1$; that is, all contrasts are associated with main effects. ■

Saturated designs are at the heart of the Taguchi quality control methods which have been attributed as being the basis of Japanese industrial supremacy. They have used these designs in developing products to determine which factors in an industrial manufacturing process affect product quality. In using these designs they are assuming that all two-factor interactions are negligible. Because of this they are not so useful in agriculture.

Resolution III designs for which *n* is a power of 2.

Rule X.3: 2^{k-p}_{III} designs in which $n = 2^{k-p}$ are constructed as follows:

1. Write down a complete design in k–p factors.
2. Compute the column of signs for additional factors by associating the factors with the interaction columns of the factors in the complete design. ■

We will now investigate the use of resolution III designs. In particular, their use in sequential experimentation will be illustrated.

Example X.10 A bike experiment

Consider the following 2^{7-4}_{III} experiment designed to investigate the effects of 7 factors on the time it takes a particular person to cycle up a hill. That is, it is a one sixteenth fraction of a 2^7 experiment involving $2^3 = 8$ runs. The design is constructed by writing down a full factorial design for the three variables **1**, **2** and **3**; the additional variables **4**, **5**, **6** and **7** are associated with all the interaction columns **12**, **13**, **23** and **123**, respectively. Hence, the generating relations of the design are **I** = **124**, **I** = **135**, **I** = **236**, **I** = **1237**. The design is shown in the following table:

A 2_{III}^{7-4} Experiment for Studying Bicycle Times

	Factor	–	+
1	Seat	up	down
2	Dynamo	off	on
3	Handlebars	up	down
4	Gear	low	medium
5	Raincoat	on	off
6	Breakfast	yes	no
7	Tyres	hard	soft

Run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	Time
1	–	–	–	+	+	+	–	69
2	+	–	–	–	–	+	+	52
3	–	+	–	–	+	–	+	60
4	+	+	–	+	–	–	–	83
5	–	–	+	+	–	–	+	71
6	+	–	+	–	+	–	–	50
7	–	+	+	–	–	+	–	59
8	+	+	+	+	+	+	+	88

Notice that all contrasts are associated with a main effect and so this design is saturated.

The experimental structure for this experiment is:

Structure	Formula
unrandomized	*8* Runs
randomized	*2* Seat+*2* Dynamo+*2* Handbars+*2* Gear +*2* Raincoat+*2* Brekkie+*2* Tyres

The pluses indicate that the factors have to be assumed to be independent.

The Genstat output file for the analysis is:

```
Genstat 5  Release 4.1  (PC/Windows NT)              10 April 2000 21:55:35
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

                 Genstat 5 Fourth Edition - (for Windows)
                 Genstat 5 Procedure Library Release PL11
                 ______________________________________________

   3  "Data taken from File: D:/ANALYSES/LM/MULTIFAC/FRF7BIKE1.GSH"
   4  DELETE [redefine=yes] Runs,Seat,Dynamo,Handbars,Gear,Raincoat,Brekkie,Tyres\
   5  ,Time
   6  FACTOR [modify=yes;nvalues=8;levels=8] Runs
   7  READ Runs; frepresentation=ordinal

    Identifier     Values   Missing     Levels
          Runs          8         0          8
```

```
FACTOR [modify=yes;nvalues=8;levels=2] Seat
READ Seat; frepresentation=ordinal

  Identifier    Values   Missing    Levels
        Seat         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Dynamo
READ Dynamo; frepresentation=ordinal

  Identifier    Values   Missing    Levels
      Dynamo         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Handbars
READ Handbars; frepresentation=ordinal

  Identifier    Values   Missing    Levels
    Handbars         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Gear
READ Gear; frepresentation=ordinal

  Identifier    Values   Missing    Levels
        Gear         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Raincoat
READ Raincoat; frepresentation=ordinal

  Identifier    Values   Missing    Levels
    Raincoat         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Brekkie
READ Brekkie; frepresentation=ordinal

  Identifier    Values   Missing    Levels
     Brekkie         8         0         2

FACTOR [modify=yes;nvalues=8;levels=2] Tyres
READ Tyres; frepresentation=ordinal

  Identifier    Values   Missing    Levels
       Tyres         8         0         2

VARIATE [nvalues=8] Time
READ Time

  Identifier   Minimum      Mean   Maximum    Values   Missing
        Time     50.00     66.50     88.00         8         0

"Load data from Frf7Bike1.gsh and analyse first fraction"
PRINT Seat,Dynamo,Handbars,Gear,Raincoat,Brekkie,Tyres,Time; \
                                                 FIELD=9; DEC=0

   Seat   Dynamo Handbars     Gear Raincoat  Brekkie    Tyres     Time
      1        1        1        2        2        2        1       69
      2        1        1        1        1        2        2       52
      1        2        1        1        2        1        2       60
      2        2        1        2        1        1        1       83
      1        1        2        2        1        1        2       71
      2        1        2        1        2        1        1       50
      1        2        2        1        1        2        1       59
      2        2        2        2        2        2        2       88

BLOCK Runs
TREAT Seat+Dynamo+Handbars+Gear+Raincoat+Brekkie+Tyres
A2PLOT [PRINT=inform,effect; STRATUM=Runs; METHOD=normal; \
        GRAPH=line] Time
```

```
40.............................................................................

***** Information summary *****

All terms orthogonal, none aliased.

***** Tables of effects *****

Variate: Time

***** Runs stratum *****

Seat Y-effect               3.50   s.e. *         rep. 4

Dynamo Y-effect            12.00   s.e. *         rep. 4

Handbars Y-effect           1.00   s.e. *         rep. 4

Gear Y-effect              22.50   s.e. *         rep. 4

Raincoat Y-effect           0.50   s.e. *         rep. 4

Brekkie Y-effect            1.00   s.e. *         rep. 4

Tyres Y-effect              2.50   s.e. *         rep. 4



             -
N            I
o            I                                                      *
r     1.0721 I
m            I                            *
a            I
l            I       *
             I
Q            I     *
u    -0.2144 I
a            I *
n            I
t            I *
i            I
l            I                                                              *
e    -1.5009 I
             -+---------+---------+---------+---------+---------+---------+-
           0.55      4.58      8.62     12.65     16.68     20.72     24.75

                                            effect
```

The estimated effects and the abbreviated aliasing pattern (ignoring interactions with three or more factors) for this experiment are as follows:

Estimated Effects and Abbreviated Aliasing Pattern

Seat	$\ell_{\mathbf{1}} = 3.5 \;\rightarrow$	**1** + **24** + **35** + **67**
Dynamo	$\ell_{\mathbf{2}} = 12.0 \;\rightarrow$	**2** + **14** + **36** + **57**
Handlebars	$\ell_{\mathbf{3}} = 1.0 \;\rightarrow$	**3** + **15** + **26** + **47**
Gear	$\ell_{\mathbf{4}} = 22.5 \;\rightarrow$	**4** + **12** + **56** + **37**
Raincoat	$\ell_{\mathbf{5}} = 0.5 \;\rightarrow$	**5** + **13** + **46** + **27**
Breakfast	$\ell_{\mathbf{6}} = 1.0 \;\rightarrow$	**6** + **23** + **45** + **17**
Tyres	$\ell_{\mathbf{7}} = 2.5 \;\rightarrow$	**7** + **34** + **25** + **16**
	$(\ell_{\mathbf{I}} = 66.5 \;\rightarrow$	average)

Suppose that previous experience suggests the standard deviation for repeated runs up the hill is about 3 seconds. Then, as each effect is the difference of two averages based on 4 observations, the standard error of the effects is about $3\sqrt{\tfrac{1}{4}+\tfrac{1}{4}} = 2.1$. Evidently, only two contrasts are distinguishable from uncontrolled variation: $\ell_{\mathbf{2}}$ and $\ell_{\mathbf{4}}$.

The interpretation of the experiment would appear to be that only factors **2** (dynamo) and **4** (gear) affect the time and they do this independently. Having the dynamo on adds about 12 seconds and using medium gear rather than low adds about 22 seconds.

However, there is some ambiguity in these conclusions. It may be, for example, that $\ell_{\mathbf{4}}$ is large, not because of a large main effect **4**, but because one or more of the interactions **12**, **56** and **37** are large. A further experiment would concentrate on factor **4** to confirm that there is a large main effect for this factor.

Also, note that, being a resolution III design, it will contain complete 2^2 factorial experiments replicated twice in every pair of factors. Thus, it contains a complete factorial in factors **2** and **4**.

Other $2_{\text{III}}^{7\text{-}4}$ *fractions*

The design we have used is only one particular 1/16 fraction of a full 2^7 design. How can the other one-sixteenth fractions be generated? The first design was generated by setting

$$\mathbf{4} = \mathbf{12},\;\; \mathbf{5} = \mathbf{13},\;\; \mathbf{6} = \mathbf{23},\;\; \mathbf{7} = \mathbf{123}$$

but, for example, we could equally have used

$$\mathbf{4} = -\mathbf{12},\;\; \mathbf{5} = \mathbf{13},\;\; \mathbf{6} = \mathbf{23},\;\; \mathbf{7} = \mathbf{123}.$$

The fraction that would be generated in this case is:

A second 2_{III}^{7-4} experiment for studying bicycle times

	Factor	–	+
1	Seat	up	down
2	Dynamo	off	on
3	Handlebars	up	down
4	Gear	low	medium
5	Raincoat	on	off
6	Breakfast	yes	no
7	Tyres	hard	soft

Run	**1**	**2**	**3**	**4** **–12**	**5** **13**	**6** **23**	**7** **123**	Time
1	–	–	–	–	+	+	–	47
2	+	–	–	+	–	+	+	74
3	–	+	–	+	+	–	+	84
4	+	+	–	–	–	–	–	62
5	–	–	+	–	–	–	+	53
6	+	–	+	+	+	–	–	78
7	–	+	+	+	–	+	–	87
8	+	+	+	–	+	+	+	60

This fraction has been obtained by reversing the signs of column **4** and this produces a completely new set of factor combinations, none having been included in the first experiment. Of course, the experimental structure remains the same.

However, the aliasing pattern has also been altered. It can be obtained from the original by reversing the sign of any interaction term involving **4** and the terms aliased with **4**. Thus, it is now:

Estimated Effects and Abbreviated Aliasing Pattern

Seat	$\ell_1' = 0.8$	➔ **1** – **24** + **35** + **67**
Dynamo	$\ell_2' = 10.2$	➔ **2** – **14** + **36** + **57**
Handlebars	$\ell_3' = 2.7$	➔ **3** + **15** + **26** – **47**
Gear	$\ell_4' = 25.2$	➔ **4** – **12** – **56** – **37**
Raincoat	$\ell_5' = -1.7$	➔ **5** + **13** – **46** + **27**
Breakfast	$\ell_6' = 2.2$	➔ **6** + **23** – **45** + **17**
Tyres	$\ell_7' = -0.7$	➔ **7** – **34** + **25** + **16**

There are 16 different ways of allocating signs to the four generators:

$$\mathbf{4} = \pm\mathbf{12},\ \ \mathbf{5} = \pm\mathbf{13},\ \ \mathbf{6} = \pm\mathbf{23},\ \ \mathbf{7} = \pm\mathbf{123}$$

Thus, switching the signs of columns **4**, **5**, **6**, and **7** will produce the 16 fractions that make up the complete factorial and the corresponding sign switching for the aliasing pattern produces the appropriate aliasing pattern.

Plackett and Burman Designs

Definition X.21: Plackett and Burman designs are two-level fractional factorial designs for studying $k = n–1$ factors in n runs where n is a multiple of 4. ■

If n is a power of 2, these are identical to those previously presented. However, the Plackett and Burman designs for $n = 12$, 20, 24 ,28 and 30 are sometimes of interest.

Example X.11: Plackett and Burman design for 11 factors in 12 runs

	Factor										
Run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**
1	+	–	+	–	–	–	+	+	+	–	+
2	+	+	–	+	–	–	–	+	+	+	–
3	–	+	+	–	+	–	–	–	+	+	+
4	+	–	+	+	–	+	–	–	–	+	+
5	+	+	–	+	+	–	+	–	–	–	+
6	+	+	+	–	+	+	–	+	–	–	–
7	–	+	+	+	–	+	+	–	+	–	–
8	–	–	+	+	+	–	+	+	–	+	–
9	–	–	–	+	+	+	–	+	+	–	+
10	+	–	–	–	+	+	+	–	+	+	–
11	–	+	–	–	–	+	+	+	–	+	+
12	–	–	–	–	–	–	–	–	–	–	–

The manner in which two-factor interactions alias main effects for most Plackett and Burman designs is complicated.

e) Table of fractional factorial designs

The following table is an extract from Table 12.15 on p. 410 of Box, Hunter and Hunter. In each cell, the type of factorial is given in the top lefthand corner. In the bottom righthand corner, either the number of times the full factorial design must be repeated or the generators for the fractional design are provided.

To use this table, examine the appropriate column if the number of factors to be investigated is fixed; examine the appropriate row if the number of runs is fixed. Note also that designs along the same leftright diagonal have the same degree of replication or fractionation.

Number of runs	Number of factors: 3	4	5	6	7	8
4	2_{III}^{3-1} **±3=12**					
8	2^3	2_{IV}^{4-1} **±4=123**	2_{III}^{5-2} **±4=12** **±5=13**	2_{III}^{6-3} **±4=12** **±5=13** **±6=23**	2_{III}^{7-4} **±4=12** **±5=13** **±6=23** **±7=123**	
16	2^3 2 times	2^4	2_{V}^{5-1} **±5=1234**	2_{IV}^{6-2} **±5=123** **±6=234**	2_{IV}^{7-3} **±5=123** **±6=234** **±7=134**	2_{IV}^{8-4} **±5=234** **±6=134** **±7=123** **±8=124**
32	2^3 4 times	2^4 2 times	2^5	2_{VI}^{6-1} **±6=12345**	2_{IV}^{7-2} **±6=1234** **±7=1245**	2_{IV}^{8-3} **±6=123** **±7=124** **±8=2345**
64	2^3 8 times	2^4 4 times	2^5 2 times	2^6	2_{VII}^{7-1} **±7=123456**	2_{V}^{8-2} **±7=1234** **±8=1256**
128	2^3 16 times	2^4 8 times	2^5 4 times	2^6 2 times	2^7	2_{VIII}^{8-1} **±8=1234567**

f) Generation of fractional factorial designs in Genstat

The Genstat command *Stats > Design > Select Design* can be used to obtain layouts for fractional factorials. The *fractional factorial designs (with blocking)* option is used with the fraction specified to be *p*. The series of questions that you will be asked and suggested responses are as follows:

How many levels do the treatments have: 2, 3?	2
Which fraction do you want: 2, 4, 8, 16, 32, 64?	*p*
How many treatment factors (3 to 12)?	*k*
Number of units per block can be either 2, 4 , …xx. The default of xx indicates no blocking. How many units do you want?	xx
What would you like to call treatment factor 1?	name for factor 1
What would you like to call treatment factor 2?	name for factor 2
.	.
What would you like to call treatment factor k?	name for factor k
Seed for randomization (0 for none)?	6-digit number
Do you want to print the design?	yes
Do you want to check the design by ANOVA?	yes

It produces just one of the fractions, not necessarily the principal fraction.

g) Computation in Genstat

Generally the analysis of fractional 2^k designs is based the normal probability plot of Yates effects, followed by an ANOVA for fitted model so that tables of means and residuals can be obtained and diagnostic checking performed.

Example X.10 A bike experiment (continued)

The following commands are used to analyse the first fraction:

```
"Load data from Frf7Bike1.gsh and analyse first fraction"
PRINT Seat,Dynamo,Handbars,Gear,Raincoat,Brekkie,Tyres,Time; \
                                                       FIELD=9; DEC=0
BLOCK Runs
TREAT Seat+Dynamo+Handbars+Gear+Raincoat+Brekkie+Tyres
A2PLOT [PRINT=inform,effect; STRATUM=Runs; METHOD=normal; \
        GRAPH=line] Time
```

Note the use of ‘+’ in the TREAT statement so that only main effects are fitted. Diagnostic checking is not performed in this case because of the very low residual degrees of freedom.