Statistics 514: Design of Experiments

Topic 11

Topic Overview

This topic will cover

- 2^k Factorial Design
- Blocking/Confounding
- Fractional Factorial Designs
- 3^k Factorial Design

2^k Factorial Design

- Each factor has two levels (often labeled + and -)
- Very useful design for preliminary analysis
- Can "weed out" unimportant factors
- Also allows initial study of interactions
- For general two-factor factorial model

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{i,j} + \epsilon_{i,j,k}$$

- Have $1 + (a-1) + (b-1) + (a-1)(b-1)$ parameters
- 2^2 factorial has only four parameters

$$\begin{array}{ccc} \alpha_1 = -\alpha_2 & & \beta_1 = -\beta_2 \\ (\alpha\beta)_{1,1} = -(\alpha\beta)_{2,2} & (\alpha\beta)_{1,2} = -(\alpha\beta)_{2,2} & (\alpha\beta)_{2,1} = -(\alpha\beta)_{2,2} \end{array}$$

- Can study 2 factors through repeated 2^2 designs
  - Select different levels of each factor

Design for 2^2 Factorial

- Label the levels of factors A and B using + and -
- Factors need not be on numeric scale
- There are 4 experimental combinations labeled

A	B	Symbol
-	-	(1)
+	-	a
-	+	b
+	+	ab

- Can express combination in terms of model parameters

$$\begin{aligned} ab = (+,+): &\quad E(y) = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{2,2} \\ a = (+,-): &\quad E(y) = \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{2,1} \\ b = (-,+): &\quad E(y) = \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{1,2} \\ (1) = (-,-): &\quad E(y) = \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{1,1} \end{aligned}$$

$$\begin{aligned} ab = (+,+): &\quad E(y) = \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{2,2} \\ a = (+,-): &\quad E(y) = \mu + \alpha_2 - \beta_2 - (\alpha\beta)_{2,2} \\ b = (-,+): &\quad E(y) = \mu - \alpha_2 + \beta_2 - (\alpha\beta)_{2,2} \\ (1) = (-,-): &\quad E(y) = \mu - \alpha_2 - \beta_2 + (\alpha\beta)_{2,2} \end{aligned}$$

Estimating Model Parameters

- Have four equations and four unknowns

$$\begin{aligned} \hat{\mu} = & & &= \frac{ab + a + b + (1)}{4n} \\ \hat{\alpha}_2 = & \quad 0.5\left(\tfrac{ab+a}{2n} - \tfrac{b+(1)}{2n}\right) & &= \frac{ab + a - b - (1)}{4n} \\ \hat{\beta}_2 = & \quad 0.5\left(\tfrac{ab+b}{2n} - \tfrac{a+(1)}{2n}\right) & &= \frac{ab - a + b - (1)}{4n} \\ (\hat{\alpha\beta})_{2,2} = & \quad 0.5\left(\tfrac{ab-b}{2n} - \tfrac{a-(1)}{2n}\right) & &= \frac{ab - a - b + (1)}{4n} \end{aligned}$$

where n represents the number of replications of each combination and ab, b, etc. represent the sample sums of each combination.

- Main effect of A defined as $2\hat{\alpha}_2$

- Main effect of B defined as $2\hat{\beta}_2$
- Interaction effect defined as $2(\hat{\alpha\beta})_{2,2}$
- Each effect function of ± 1 (combination sums)

Using Effects to Draw Inference

- Look at magnitude and direction of effects
- ANOVA then used to verify conclusions (Chapter 5)
- Effects and contrasts related as shown

	Contrast coefficients			
Effect	(1)	a	b	ab
A	-1	+1	-1	+1
B	-1	-1	+1	+1
AB	1	-1	-1	+1

- AB coefficients simply product of A and B coefficients
- Sum of squares related to model effects

$$SS_{Contrast} = \frac{(\sum c_i y_i)^2}{n \sum c_i^2} \qquad i = 1,2,3,4$$

$$SS_A = \frac{(ab+a-b-(1))^2}{4n} = 4n\hat{\alpha}_2^2$$

Regression/Response Surface

- Code each factor as ± 1 where $+1$ is high level
- x_1 associated with Factor 1 $= A$
- x_2 associated with Factor 2 $= B$
- Regression Model

$$y_{i,j,k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2 + \epsilon_{i,j,k}$$

- Can show

$$\beta_0 = \hat{\mu} \qquad \beta_1 = \hat{\alpha}_2 \qquad \beta_2 = \hat{\beta}_2$$
$$\beta_{1,2} = (\hat{\alpha\beta})_{2,2}$$

- Parameters based on unit increase (not -1 to +1)
- Can use regression model to form contours
- No interaction $\rightarrow$ additive relationship
- Can often see direction for improvement in response

Hormone Example

Combination	Total
(1)	467
a	394
b	642
ab	571

$$
\begin{aligned}
\hat{\mu} &= \tfrac{571+394+642+467}{4(6)} &= 90.75 \\
\hat{\alpha}_2 &= \tfrac{571+394-642-467}{4(6)} &= -6 \\
\hat{\beta}_2 &= \tfrac{571-394+642-467}{4(6)} &= 14.6666 \\
(\hat{\alpha\beta}_{2,2}) &= \tfrac{572-394-642+467}{4(6)} &= 0.083333
\end{aligned}
$$

- $SS_A = 4(6)\hat{\alpha}_2^2 = 864$
- $SS_B = 4(6)\hat{\beta}_2^2 = 5162.67$
- $SS_{AB} = 4(6)\hat{\alpha\beta}_{2,2}^2 = 0.17$
- Response Surface model

$$y = 90.75 - 6x_1 + 14.67x_2 + \epsilon$$

To increase response, model says to increase level. Be careful of interpretation.

2^3 Factorial Designs

- Eight combinations displayed as cube

A	B	C	Symbol
-	-	-	(1)
+	-	-	a
-	+	-	b
+	+	-	ab
-	-	+	c
+	-	+	ac
-	+	+	bc
+	+	+	abc

- A main effect is equal to

$$
\begin{gathered}
2\hat{\alpha}_2 \\
\tfrac{(a-(1))+(ab-b)+(ac-c)+(abc-bc)}{4n} \\
\tfrac{(a-(1))(b+(1))(c+(1))}{4n}
\end{gathered}
$$

- The AB interaction is equal to

$$2(\hat{\alpha\beta})_{2,2}$$
$$\frac{((abc-bc)+(ab-b))-((ac-c)+(a-(1)))}{4n}$$
$$\frac{(a-(1))(b-(1))(c+(1))}{4n}$$

2^3 Factorial

	Contrast Coefficients						
Symbol	A	B	AB	C	AC	BC	ABC
(1)	-	-	+	-	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	+	-	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	-	+	+	-	-
bc	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+

$$\begin{aligned}\text{Contrast} &= \sum c_i y_i = (\pm(1) \pm a \pm b \pm ab \pm c \pm ac \pm bc \pm abc)\\ \text{Parameter} &= \text{Contrast}/8n\\ \text{Effect} &= \text{Contrast}/4n = 2\text{ Parameter}\end{aligned}$$

- $SS_{Effect} = (\text{Contrast})^2/8n$

- $SS_{Effect} = 2n(\text{Effect})^2$

- $\text{Var}(\text{Contrast}) = 8n\sigma^2$
- $\text{Var}(\text{Effect}) = \sigma^2/2n$

General 2^k Factorial

- Contrast is sum of treatment combinations (± 1 coefficients)

$$\text{Var}(\text{Contrast}) = n2^k\sigma^2$$

- Effect is Contrast divided by $n2^{k-1}$

$$\text{Var}(\text{Effect}) = \sigma^2/n2^{k-2}$$

- Can use variances to test significance of effect

Approximate 95% CI for Effect
$$\text{Effect } \pm 1.96\sqrt{\sigma^2/n2^{k-2}}$$

- Look at effects whose intervals do not contain 0

Yates Algorithm

- For general 2^k design
  - List data in *standard order*
  - Create k new columns in the following manner
  - The first 2^{k-1} entries are obtained by adding adjacent values
  - The last 2^{k-1} entries are obtained by subtracting adjacent values
  - Estimates of effects obtained by dividing kth column by $n2^{k-1}$
  - Sum of Squares obtained by squaring kth column and dividing by $n2^k$

Combination	y	1	2	3	Estimates	SS
(1)	15	25	55	110	27.5	1512.5
a	10	30	55	20	5	50
b	5	40	15	-20	-5	50
ab	25	15	5	10	2.5	12.5
c	15	-5	5	0	0	0
ac	25	20	-25	-10	-2.5	12.5
bc	10	10	25	-30	-7.5	112.5
abc	5	-5	-15	-40	-10	200

Single Replicate in 2^k Design

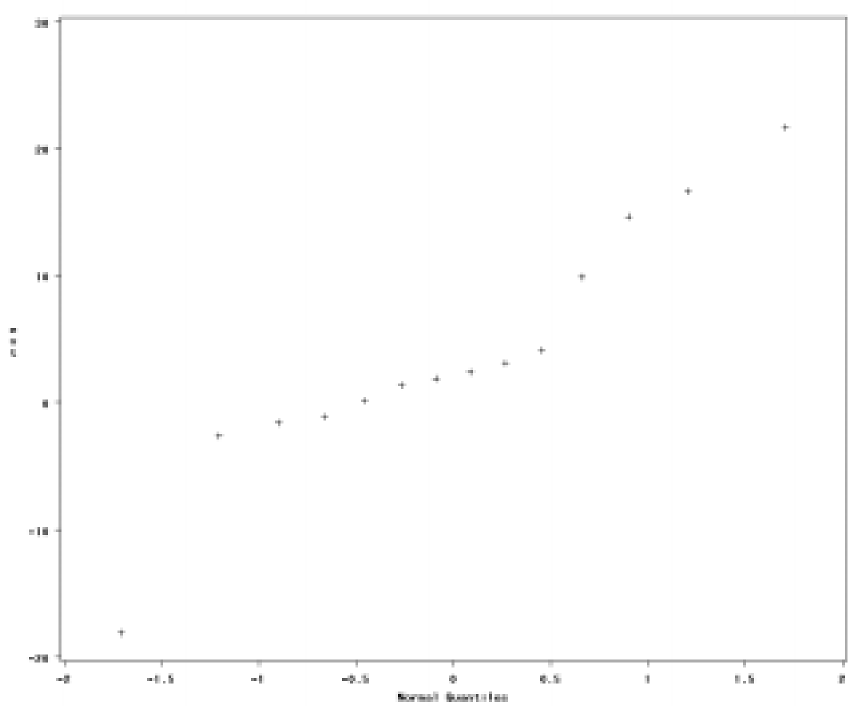
- With only one replicate
  - Cannot estimate all interactions and σ^2
- Often assume 3 or higher interactions do not occur
- Pool estimate together for error
- Warning: may pool significant interaction

- Create Normal Probability Plot of Effects
- If negligible, effect will be $N(0, \sigma^2)$
- Will fall along straight line
- Significant effects will not lie along same line
- Use Yates algorithm or SAS to create plot

Example 6-2

```
data new;
  input fact1 fact2 fact3 fact4 est;
  cards;
  2 1 1 1  21.625
  1 2 1 1   3.125
  2 2 1 1   0.125
  1 1 2 1   9.875
  2 1 2 1 -18.125
  1 2 2 1   2.375
  2 2 2 1   1.875
  1 1 1 2  14.625
  1 2 1 2  16.625
  2 2 1 2   4.125
  1 1 2 2  -1.125
  2 1 2 2  -1.625
  1 2 2 2  -2.625
  2 2 2 2   1.375
;

proc univariate noprint;
  qqplot est/normal;
```




```
data filter;
  do D = -1 to 1 by 2; do C = -1 to 1 by 2;
```

```
  do B = -1 to 1 by 2; do A = -1 to 1 by 2;
  input y @@;  output;
  end; end; end; end;
cards;
45 71 48 65 68 60 80 65 43 100 45 104 75 86 70 96
;

data inter;                       /* Define Interaction Terms */
  set filter;
  AB = A*B; AC = A*C; AD = A*D; BC = B*C; BD = B*D; CD = C*D; ABC=AB*C; ABD = AB*D;
  ACD = AC*D; BCD = BC*D; ABCD = ABC*D;

proc glm data = inter;            /* GLM Proc to Obtain Effects */
  class A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;
  model y = A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;
  estimate 'A' A 1 -1; estimate 'AC' AC 1 -1;

proc reg outest = effects data = inter;  /* REG Proc to Obtain Effects */
  model y = A B C D AB AC AD BC BD CD ABC ABD ACD BCD ABCD;
data effect2;  set effects;
  drop y intercept _RMSE_;
proc transpose data = effect2 out = effect3;
data effect4; set effect3; effect = col1*2;
proc sort data = effect4; by effect;
proc print data = effect4;
proc rank data = effect4 normal = blom;
  var effect;  ranks neff;

symbol1 v = circle;
proc gplot;
  plot neff*effect = _NAME_;
```

Adding Center Points to 2^k Design

- Suppose factor effects continuous
- Suppose instead of

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{i,j} x_i x_j + \epsilon$$

the model is instead

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{i,j} x_i x_j + \sum_{j=1}^{k} \beta'_j x_j^2 + \epsilon,$$

where β'_j are *quadratic effects*.

If we add points in the center of the design $(0,0)$

- can estimate quadratic effects
- estimates of effects are the same. Why?

$$SS_{quad} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C},$$

  where

  $\bar{y}_F$ – average of cell means at factor level
  $\bar{y}_C$ – average of cells at center point

By adding other center points (e.g., $(0,0,-1)$ and $(0,0,1)$ in addition to $(0,0,0)$), get other estimates of curvature (e.g., $\beta_{i,j,.} x_i x_j^2$).

Fundamental Principles for Factorial Effects

Hierarchical Ordering Principle

- Lower order effects are more likely to be important than higher order effects
- Effects of the same order are equally likely to be important.

Effect Sparsity Principle

The number of relatively important effects in a factorial experiment is small. True model is often (close to) linear.

Effect Heredity Principle

In order for an interaction to be significant, at least one of its parent factors should be significant.

Tests are not designed for principles.

Miscellaneous

- Replicates vs repetition: combinations of factors are often not assigned completely randomly
- Folding in error – if the effect **and all interactions with that effect** are 0, fold in the factor, particularly if there are no replicates.
- Transformations sometimes get rid of interactions.
- Yates algorithm – $k2^k$ vs 2^{2k} steps

Blocking in 2^k Factorial Designs

- For RCBD, each combination run in each block
  - $2^2 \rightarrow 4$ EU's per block
  - $2^3 \rightarrow 8$ EU's per block
  - Randomize run order within block
- Suppose you cannot run all combinations within block
- Must do some sort of incomplete block analysis
- If you do not, certain effects confounded
- Confounding: two effects are indistinguishable
- May "sacrifice" certain effects thought to be small
- 2^k design makes set-up simple

Confounding in 2^k with only 2 blocks

- Blocks assumed to allow 2^{k-1} combinations
- First consider 2^2 factorial (2 combinations per block)
- Possible pairings
  1. (1) and b together $\rightarrow$ a and ab together
  2. (1) and a together $\rightarrow$ b and ab together
  3. (1) and ab together $\rightarrow$ a and b together

1. Effect of A, $(ab + a - b - (1))/2$, is block difference
2. Effect of B, $(ab - a + b - (1))/2$, is block difference
   Both have a main effect confounded with block
3. Effect of AB, $(ab - a - b + (1))/2$, is block difference
   Allows for main effect estimates (blocks cancel out).

Choice of Confounding Factors

- Common to confound highest order interaction
- Can use +/− table to determine blocks
- For two factor, recall the following table

A	B	AB	Symbol
-	-	+	(1)
+	-	-	a
-	+	-	b
+	+	+	ab

- Use confounding column to determine blocks

  +'s in block 1 and -'s in block 2

- Consider three factor

A	B	C	AB	AC	BC	ABC	Symbol
-	-	-	+	+	+	-	(1)
+	-	-	-	-	+	+	a
-	+	-	-	+	-	+	b
+	+	-	+	-	-	-	ab
-	-	+	+	-	-	+	c
+	-	+	-	+	-	-	ac
-	+	+	-	-	+	-	bc
+	+	+	+	+	+	+	abc

  – Best assignment would be a, b, c, abc together
  – Can estimate all but three factor interaction

2^k Factorial in Four Blocks

- Four blocks each containing 2^{k-2} EU's
- Useful in situations where $k \geq 4$
- Must select two effects to confound
- Will result in a third confounded factor
- Consider 6 factor factorial run in 4 blocks of 16 EU's
  – Block 1 uses ABC + and DEF +
  – Block 2 uses ABC + and DEF -
  – Block 3 uses ABC - and DEF +
  – Block 4 uses ABC - and DEF -
- Results in $(ABC)(DEF) = ABCDEF$ confounded
  – $ABCD$ and $DEF \rightarrow ABCEF$ confounded
  – AB and $ABEF \rightarrow EF$ confounded

- Can extend to 8 and 16 blocks
- Table 7-9 summarizes these designs (page 277)

Partial Confounding

- Can replicate blocking design
- Confound different effects each replication
- Allows estimation of all effects
  - Confounded effects based on nonconfounded replicates
  - Use Yates' Algorithm for all nonconfounded effects
  - See Example 7-3 (page 279)

```
/* Example 7-3 */
data cool;
input block fact1 fact2 fact3 y;
cards;
  1 -1 -1 -1 -3
  1  1  1 -1  2
  1  1 -1  1  2
  1 -1  1  1  1
  2  1 -1 -1  0
  2 -1  1 -1 -1
  2 -1 -1  1 -1
  2  1  1  1  6
  3 -1 -1 -1 -1
  3 -1 -1  1  0
  3  1  1 -1  3
  3  1  1  1  5
  4  1 -1 -1  1
  4 -1  1 -1  0
  4  1 -1  1  1
  4 -1  1  1  1
;

proc glm;
  class fact1 fact2 fact3 block;
  model y = block fact1|fact2|fact3;
```

```
Dependent Variable: y
                                          Sum of
      Source                      DF         Squares     Mean Square    F Value    Pr > F
      Model                       10     74.25000000      7.42500000       9.90    0.0103
      Error                        5      3.75000000      0.75000000
      Corrected Total             15     78.00000000
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	3	3.50000000	1.16666667	1.56	0.3101
fact1	1	36.00000000	36.00000000	48.00	0.0010
fact2	1	20.25000000	20.25000000	27.00	0.0035
fact1*fact2	1	0.50000000	0.50000000	0.67	0.4513
fact3	1	12.25000000	12.25000000	16.33	0.0099
fact1*fact3	1	0.25000000	0.25000000	0.33	0.5887
fact2*fact3	1	1.00000000	1.00000000	1.33	0.3004
fact1*fact2*fact3	1	0.50000000	0.50000000	0.67	0.4513

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	3	1.25000000	0.41666667	0.56	0.6667
fact1	1	36.00000000	36.00000000	48.00	0.0010
fact2	1	20.25000000	20.25000000	27.00	0.0035
fact1*fact2	1	0.50000000	0.50000000	0.67	0.4513
fact3	1	12.25000000	12.25000000	16.33	0.0099
fact1*fact3	1	0.25000000	0.25000000	0.33	0.5887
fact2*fact3	1	1.00000000	1.00000000	1.33	0.3004
fact1*fact2*fact3	1	0.50000000	0.50000000	0.67	0.4513

Example

Consider a 2^3 factorial run in 4 blocks

Each replicate will result in 3 confounded effects

Consider 4 replicates for 32 total observations
Replicate 1: Confound BC and $AC \rightarrow AB$

Replicate 2: Confound BC and $ABC \rightarrow A$

Replicate 3: Confound AC and $ABC \rightarrow B$

Replicate 4: Confound AB and $ABC \rightarrow C$

Three replicates to estimate A, B, and C

Two replicates to estimate AB, AC, and BC

One replicate to estimate ABC

Data

Replicate 1 – AB, AC, and BC confounded

Blk 1	Blk 2	Blk 3	Blk 4
75 (1)	89 *ab*	61 *a*	30 *b*
100 *abc*	73 *c*	45 *bc*	54 *ac*

Replicate 2 – *A*, *BC*, *ABC* confounded

Blk 1	Blk 2	Blk 3	Blk 4
60 (1)	47 *a*	1 *b*	26 *ac*
34 *bc*	81 *abc*	35 *c*	52 *ab*

Replicate 3 – *B*, *AC*, *ABC* confounded

Blk 1	Blk 2	Blk 3	Blk 4
58 (1)	48 *a*	18 *b*	68 *ab*
42 *ac*	52 *c*	82 *abc*	32 *bc*

Replicate 4 – *C*, *AB*, *ABC* confounded

Blk 1	Blk 2	Blk 3	Blk 4
47 (1)	34 *a*	50 *c*	37 *ac*
57 *ab*	4 *b*	80 *abc*	27 *bc*

Since no effect is estimated from all replications (except error), we will compute the effect sums associated with each replicate and combine the appropriate information.

The following are the column 3 sums using Yates' Algorithm for each of the replicates. Only the sums that are not confounded with the particular replicate are presented

Effect	Rep 1	Rep 2	Rep 3	Rep 4
(1)	527	336	400	336
A	**81**	-	**80**	**80**
B	**1**	**0**	-	**0**
AB	-	**120**	**120**	-
C	**17**	**16**	**16**	-
AC	-	**0**	-	**0**
BC	-	-	**40**	**40**
ABC	**1**	-	-	-

$$
\begin{aligned}
SS_A &= \frac{(81+80+80)^2}{3(8)} \\
SS_B &= \frac{(1+0+0)^2}{3(8)} \\
SS_{AB} &= \frac{(120+120)^2}{2(8)} \\
SS_C &= \frac{(16+16+17)^2}{3(8)} \\
SS_{AC} &= \frac{(0+0)^2}{2(8)} \\
SS_{BC} &= \frac{(40+40)^2}{2(8)} \\
SS_{ABC} &= \frac{1^2}{1(8)}
\end{aligned}
$$

Using SAS

```
options nocenter ps = 50 ls = 80;

data new;
  input repl blk a b c resp;
  cards;
  1 1 0 0 0 75
  1 1 1 1 1 100
  1 2 1 1 0 89
  1 2 0 0 1 73
  1 3 1 0 0 61
  1 3 0 1 1 45
  1 4 0 1 0 30
  1 4 1 0 1 54
  2 1 0 0 0 60
  2 1 0 1 1 34
  2 2 1 0 0 47
  2 2 1 1 1 81
  2 3 0 1 0 1
  2 3 0 0 1 35
  2 4 1 0 1 26
  2 4 1 1 0 52
  3 1 0 0 0 58
  3 1 1 0 1 42
  3 2 1 0 0 48
  3 2 0 0 1 52
  3 3 0 1 0 18
```

```
  3 3 1 1 1 82
  3 4 1 1 0 68
  3 4 0 1 1 32
  4 1 0 0 0 47
  4 1 1 1 0 57
  4 2 1 0 0 34
  4 2 0 1 0 4
  4 3 0 0 1 50
  4 3 1 1 1 80
  4 4 1 0 1 37
  4 4 0 1 1 27
;

proc glm;
  class repl blk a b c;
  model resp = repl blk(repl) a|b|c;
```

SAS Output

```
Dependent Variable: resp
                                      Sum of
Source                      DF       Squares    Mean Square   F Value   Pr > F
Model                       22   17128.71875      778.57813   28028.8   <.0001
Error                        9       0.25000        0.02778
Corrected Total             31   17128.96875

Source                      DF     Type I SS    Mean Square   F Value   Pr > F
repl                         3   3040.093750    1013.364583   36481.1   <.0001
blk(repl)                   12   7568.375000     630.697917   22705.1   <.0001
a                            1   2420.041667    2420.041667   87121.5   <.0001
b                            1      0.041667       0.041667      1.50   0.2518
a*b                          1   3600.000000    3600.000000    129600   <.0001
c                            1    100.041667     100.041667   3601.50   <.0001
a*c                          1      0.000000       0.000000      0.00   1.0000
b*c                          1    400.000000     400.000000   14400.0   <.0001
a*b*c                        1      0.125000       0.125000      4.50   0.0629


Source                      DF   Type III SS    Mean Square   F Value   Pr > F
repl                         3   3040.093750    1013.364583   36481.1   <.0001
blk(repl)                   12    653.906250      54.492188   1961.72   <.0001
a                            1   2420.041667    2420.041667   87121.5   <.0001
b                            1      0.041667       0.041667      1.50   0.2518
a*b                          1   3600.000000    3600.000000    129600   <.0001
c                            1    100.041667     100.041667   3601.50   <.0001
a*c                          1      0.000000       0.000000      0.00   1.0000
b*c                          1    400.000000     400.000000   14400.0   <.0001
a*b*c                        1      0.125000       0.125000      4.50   0.0629
```

Fractional Factorials

- May not have resources for complete factorial design
- Number of runs required for factorial grows quickly
  - Consider 2^k design
  - If $k = 7 \rightarrow 128$ runs required
  - Can estimate 127 effects
  - Only 7 df for main effects
  - 99 df are for interactions of order ≥ 3.
- Often system driven by low order effects
- Would like design
  - To utilize this sparsity principle
  - Can be projected into larger designs.
- Fractional factorials provide these options.

Example

- Suppose you were designing a new car for mileage
- Wanted to consider (2 levels each)
  - Engine Size
  - Number of cylinders
  - Drag
  - Weight
  - Automatic vs Manual
  - Shape
  - Tires
  - Suspension
  - Gas Tank Size
- Only have resources for 2^7 design

  If you drop two factors for a complete factorial, could discard significant main effects or lower order interactions

  Want option to keep all nine factors in model

Must assume higher order interactions insignificant

Are six and seven order interactions meaningful?

Two-Level Fractional Factorials

- Assume certain higher order interactions negligible
- Can then collect more info on lower level effects
- Example: Latin Square (2^3 factorial)
  - Let A = Block Factor 1, B = Block Factor 2, C = Treatment

A	B	C	AB	AC	BC	ABC	Symbol
-	-	-	+	+	+	-	(1)
+	-	-	-	-	+	+	a
-	+	-	-	+	-	+	b
+	+	-	+	-	-	-	ab
-	-	+	+	-	-	+	c
+	-	+	-	+	-	-	ac
-	+	+	-	-	+	-	bc
+	+	+	+	+	+	+	abc

  - Utilize only four observations instead of eight

A \ B	-	+
-	+	-
+	-	+

A \ B	-	+
-	-	+
+	+	-

  - Observed combinations associated with ABC column

1. May observe C, B, A, ABC
2. May observe (1), BC, AC, and AB

- Since only four observations, certain factors confounded
- Use association column to determine confounding
- Association column known as *generator*
- Use generator in the following manner

$$I = ABC \rightarrow A = BC, B = AC, \text{ and } C = AB$$

- Thus, if additivity, can estimate A, B, and C

Assumptions and Expectations

- Used when
  - Runs expensive and variance estimates available
  - Screening experiments when many factors considered
  - Sequential analysis possible (put fractions together)
- Interest in main effects and low order interactions
- Prepared to assume certain interactions are zero
- Not primarily interested in estimate of variance

Emphasis is on finding small experiments in which a high percentage of df are used for estimation of low order effects

- Notation
  - Full factorial is 2^k
  - Fractional Factorial is 2^{k-p}
  - Degree of fraction is 2^{-p}

Half-Fraction 2^k Factorials

- This is one-half the usual number of runs
- Similar to blocking procedure
  - Choose a **generator** which divides effects into two
  - Based on pluses and minuses of one factor
  - **Defining Relation:** $I =$ generator
- Consider three factor but only 4 runs possible

A	B	C	AB	AC	BC	ABC	Symbol
+	-	-	-	-	+	+	a
-	+	-	-	+	-	+	b
-	-	+	+	-	-	+	c
+	+	+	+	+	+	+	abc

- Select $I = ABC$, get groups (a, b, c, abc) and $(ab, ac, bc, (1))$
- Select $I = A$, get groups (a, ab, ac, abc) and $((1), b, c, bc)$
- Use generator to determine confounded effects

Half-Fraction 2^k Factorials

- Consider $I = ABC$ and the group (a, b, c, abc)

A	B	C	AB	AC	BC	ABC	Symbol
-	-	-	+	+	+	-	(1)
+	-	-	-	-	+	+	a
-	+	-	-	+	-	+	b
+	+	-	+	-	-	-	ab
-	-	+	+	-	-	+	c
+	-	+	-	+	-	-	ac
-	+	+	-	-	+	-	bc
+	+	+	+	+	+	+	abc

- Each effect is estimated as follows:

$$\begin{array}{rr} l_A = 0.5(a - b - c + abc) & l_{BC} = 0.5(a - b - c + abc) \\ l_B = 0.5(-a + b - c + abc) & l_{AC} = 0.5(-a + b - c + abc) \\ l_C = 0.5(-a - b + c + abc) & l_{AB} = 0.5(-a - b + c + abc) \end{array}$$

- Cannot differentiate between

  A and BC, B and AC, and C and AB

- Can use defining relation to determine confounding
- "Multiply" each side by an effect

$$\begin{array}{r} (A)I = A = (A)ABC = A^2BC = BC \\ (AB)I = AB = (AB)ABC = A^2B^2C = C \end{array}$$

- Linear combinations estimate

  A + BC, B + AC, and C + AB

Half-Fraction 2^k Factorials

- Suppose we use other group ($I = -ABC$)

$$\begin{array}{rr} l_A = 0.5(ab + ac - bc - (1)) & l_{BC} = 0.5(-ab - ac + bc + (1)) \\ l_B = 0.5(ab - ac + bc - (1)) & l_{AC} = 0.5(-ab + ac - bc + (1)) \\ l_C = 0.5(-ab + ac + bc - (1)) & l_{AB} = 0.5(ab - ac - bc + (1)) \end{array}$$

- $l_A = -l_{BC}$ so combination is estimating $A - BC$
- If both fractions were run,

- Could separately estimate effects
- ABC is confounded with blocks

- Similar to factorial in two blocks

Can piece together fractional factorial into bigger design if appropriate

Another Example

Consider a 2^{5-1} fractional factorial. Usually would have 32 runs, but we will have 16. Let us use the defining relation $I = ABCDE$. This means we will have the following confounding effects:

A and BCDE	AB and CDE	ACD and BE
AC and BDE	AD and BCE	ACE and BD
AE and BCD	ABC and DE	ABCD and E
ABD and CE	ABE and CD	ABDE and C
ADE and BC	ABCE and D	ACDE and B

Main effects confounded with third order interactions, first order interactions confounded with second order interactions

Known as a Resolution V Design

Using Yates' Algorithm

- For 2^{k-1} design, will set up $k-1$ columns
- Select $k-1$ factors and write in standard order
- Multiply columns to determine sign of last factor
- See Table 8-5 for 2^{5-1} example
- Consider these results from previous example

Combination	y	1	2	3	4	Effect
e	15	25	55	110	185	
a	10	30	55	75	35	A+BCDE
b	5	40	30	20	-35	B+ACDE
abe	25	15	45	15	15	AB+CDE
c	15	15	15	-20	15	C+ABDE
ace	25	15	5	-15	-15	AC+BDE
bce	10	30	10	10	-45	BC+ADE
abc	5	15	5	5	-35	ABC+DE
d	5	-5	5	0	-35	D+ABCE
ade	10	20	-25	15	-5	AD+BCE
bde	5	10	0	-10	5	BD+ACE
abd	10	-5	-15	-5	-5	ABD+CE
cde	15	5	25	-30	15	CD+ABE
acd	15	5	-15	-15	5	ACD+BE
bcd	5	0	0	-40	15	BCD+AE
abcde	10	5	5	5	45	ABCD+E

Resolution

- A design of resolution R is one in which no p-factor effect is confounded with any other effect containing less than $R - p$ factors
  - Resolution III design does not confound main effects with other main effects.
  - Resolution IV design does not confound main effects with two-factor interactions but does confound two-factor interactions with other two-factor interactions.
  - Resolution V design does not confound any main effects and two factor interactions with each other.
  - Resolution III – can estimate main effects if you assume no interaction
  - Resolution IV – can estimate main effects without assuming two-factor interactions negligible
  - Resolution V – can estimate main and two-factor interactions if assume higher order terms negligible
- Resolution also defined by length of shortest word in defining relation
  - $I = ABC$ is Resolution III
  - $I = ABCD$ is Resolution IV
  - $I = ABCDE$ is Resolution V

Construction of a 2^{k-1} fractional factorial with highest resolution

1. Write a full factorial design for the first $k - 1$ variables

2. Associate the kth variable with $\pm$ interaction of $k-1$

- Consider 2^{4-1} design

A	B	C	$D=ABC$	Effect	Combination
-	-	-	-	(1)	(1)
+	-	-	+	a	ad
-	+	-	+	b	bd
+	+	-	-	ab	ab
-	-	+	+	c	cd
+	-	+	-	ac	ac
-	+	+	-	bc	bc
+	+	+	+	abc	abcd

- $I = ABCD \rightarrow$ Resolution 4 design

Alternative View of Half Fraction Design

- Full factorial in $k-q$ factors embedded in 2^{k-q} design
- Consider any 2^{k-1} design
- If we collapse design by omitting variable
- Remaining is a $2^{k'}$ full factorial ($k' = k-1$)
- Can be shown for resolution R, complete factorial for $R-1$ factors
  - Consider 2^{3-1} design (Resolution III)
  - Can view combinations on cube
  - Regardless of direction, you can squeeze cube into square with observations at each corner

Another Example

- Consider the following 2^3 design

Combination	y	1	2	3	Estimate
(1)	5	17	43	81	
a	12	26	38	21	5.25
b	10	13	13	21	5.25
ab	16	25	8	-3	-0.75
c	4	7	9	-5	-1.25
ac	9	6	12	-5	-1.25
bc	11	5	-1	3	0.75
abc	14	3	-2	-1	-0.25

- Consider instead a half-fraction with $I = ABC$

A	B	$C = AB$	Combination	y	1	2	Estimate	Effect
-	-	+	c	4	16	40		
+	-	-	a	12	24	12	6.0	$A + BC$
-	+	-	b	10	8	8	4.0	$B + AC$
+	+	+	abc	14	4	-4	-2.0	$C + AB$

- Notice effects are sums of previous estimates
- Could ignore C and this becomes full factorial

Constructing Fractional Factorial

1. Choose q generators and get aliases of I.
2. Find a set of $k - q$ base factors that has an embedded complete factorial.
3. Write the factor-level combinations of the base factors in standard order.
4. Find the aliases of the remaining q factors in terms of interactions of the $k - q$ base factors.
5. Determine the plus/minus pattern for the q remaining factors from their aliased interactions.
6. Add letters to the factor-level combinations of the base factors to indicate when the remaining factors are at their high levels (plus).

Example of a 2^{8-4} Design (page 478, Oehlert)

Consider 2^{8-4} with generators $BCDE$, $ACDF$, $ABDG$, and $-ABCH$. (Aliases of I are $BCDE$, $ACDF$, $ABEF$, $ABDG$, $ACEG$, $BCFG$, $DEFG$, $-ABCH$, $-ADEH$, $-BDFH$, $-CEFH$, $-CDGH$, $-BEGH$, $-AFGH$, $-ABCDEFGH$.)

- Embedded factorial in A, B, C, and D. (Stripping away E, F, G, and H, get aliases of I are BCD, ACD, AB, ABD, AC, BC, D, $-ABC$, $-AD$, $-BD$, $-C$, $-CD$, $-B$, $-A$, $-ABCD$.)
- $E = BCD$, $F = ACD$, $G = ABD$, and $H = -ABC$.

Embedded Design	$E = BCD$	$F = ACD$	$G = ABD$	$H = -ABC$	Final Design
(1)	-1	-1	-1	1	h
a	-1	1	1	-1	afg
b	1	-1	1	-1	beg
ab	1	1	-1	1	abefh
c	1	1	-1	-1	cef
ac	1	-1	1	1	acegh
bc	-1	1	1	1	bcfgh
abc	-1	-1	-1	-1	abc
d	1	1	1	1	defgh
ad	1	-1	-1	-1	ade
bd	-1	1	-1	-1	bdf
abd	-1	-1	1	1	abdgh
cd	-1	-1	1	-1	cdg
acd	-1	1	-1	1	acdfh
abd	1	-1	-1	1	bcdeh
abcd	1	1	1	-1	abcdefg

Resolution

Design of Resolution R has R letters in shortest alias of I. (Example of notation: 2^{8-4}_{IV} is 2^{8-4} design of Resolution IV.)

Minimum Aberration

Number of short aliases

Example

$$I = ABCF = BCDG = ADCF$$
$$I = ABCF = ADEG = BCDEG$$
$$I = ABCDF = ABCEG = DEFG$$

All are Resolution IV, but third aliasing is preferred because only 1 four-letter alias.

Analyzing 2^{k-q} Design

- There is no estimate of pure error.
- Need pooling or graphical method.
- If interaction effect looks large, may be aliased effect.
- If certain effects not significant, *project* into full factorial on smaller dimension.

Example

2^{7-2} design with $I = ABCDF = ABCEG = DEFG$

- Contains a replicated factorial in any set of 3 factors
- Contains complete factorial in all sets of 4 factors, except $DEFG$.

- Can make p-values too optimistic

Troubles with De-aliasing

Don't get everything for free.

Consider a 2^{5-2} design with

$$I = ABC = -CDE = -ABDE$$

- Suppose A, C, E look significant

$$\begin{aligned} A &= BC = -ACDE = BDE \\ C &= AB = -DE = -ABCDE \\ E &= ABCE = -CD = -ABD \end{aligned}$$

  Easy to believe A, C, and E significant.

- Suppose A, B, C look significant

$$\begin{aligned} A &= BC = -ACDE = -BDE \\ B &= AC = -BCDE = -ADE \\ C &= AB = -DE = -ABCDE \end{aligned}$$

  Is it

  A, B, and C?
  A, B, and AB?
  B, C, and BC?
  A, C, and AC?

  Need external info or additional data

Sequential Use of Factorial Designs

- Often more efficient to look at half fraction
- Analyze results
- Decide on best set of runs for next experiment

Can add or remove factors
Change responses
Vary factors over new ranges

- If ambiguities, can run remaining half of factorial
- Only lose estimate of highest order interaction
- Important to always randomize order of runs
- Can use `proc factex` to generate design
- Need quick turnaround

The General 2^{k-p} Fractional Factorial

- Must select p independent generators
- Want to have best alias relationships
- Defining relation based on $2^p - 1$ effects
- Often try to maximize the resolution
- Each effect has $2^p - 1$ aliases
- Often assume higher order interactions zero
- Simplifies alias structure
- Table 8-14 summarizes potential generators
  - $k \leq 15$ and $n \leq 128$
  - Results in highest possible resolution
- Table X: alias relationships for designs with $n \leq 64$
- Can use Yates' in similar fashion to obtain estimates

Example

Consider 2^{5-2} which consists of eight runs. Suppose we choose $I = ABC$ and $I = BDE$. The defining relation is $I = ABC = BDE = ACDE$ so this was resolution III design. The factor A is aliased with BC, $ABDE$, and CDE.

Consider 2^{11-4} which has 128 runs. We start with a complete 2^7. Consider the factors A, B, C, D, E, G, J with remaining factors F, H, K, L. Defines four generators as

- $F = ABCDE$

- $K = ABFJ$
- $L = AEFGK$
- $H = ACEL$

The defining relation is $I = ABCDEF = ABFJK = AEFGKL = ACEHL$. If multiply these together in pairs, we get $I = CDEJK = BCDGKL = BDFHL = BEGJL = BCEFJHKL = CFGHK$. If we multiply these in triples, we get $I = ABCGHK = ABDEGHK = ACDFG = ADHJL$, and if we multiply all four together, we get $I = DEFGHK$. Since the shortest word in relation is of length five, this has resolution V.

The General 2^{k-p} Fractional Factorial

- The 2^{k-p} collapses into either a
  - Full factorial
  - Fractional factorial of subset $r \leq k - p$
- Can block fractional factorials if necessary
  - Presented in Table X
  - Minimum block size for designs is of size 8
  - Block to confound high order interaction
- Blocking may change resolution of design

Resolution III Designs

- Can use this design to efficiently investigate numerous factors
- Can use Resolution III to investigate $N - 1$ factors in N runs
- N must be a multiple of 4
  1. 4 runs to investigate 3 factors $\rightarrow 2^{3-1}$ design
  2. 8 runs to investigate 7 factors $\rightarrow 2^{7-4}$ design
  3. 16 runs to investigate 15 factors $\rightarrow 2^{15-11}$ design
- For 2^{3-1} design
  - Each main effect aliased with one two-factor interaction
  - Introduced fractional factorials with design
- For 2^{7-4} design
  - Each main effect aliased with three two-factor interactions

- Often ignore interactions of order ≥ 3

- For 2^{15-11} design
  - Each main effect aliased with seven two-factor interactions

Sequential Assembly of Fractions

- Consider 2^{7-4} design with $D = AB$, $E = AC$, $F = BC$, and $G = ABC$
- If factor D important and don't want it confounded
  - Use same generators except $D = -AB$
  - If all three factor interactions zeros, can estimate D
  - Can also estimate all interactions concerning D
- Can use two 2^{7-4} to get resolution IV
  - Instead of flipping one sign, flip sign of all factors
  - Generators of even size flip sign
  - Known as *folding over*
  - Breaks link between main effects and two-factor interactions
- $D = AB$, $E = AC$, $F = BC$, and $G = ABC$
- $D = -AB$, $E = -AC$, $F = -BC$, and $G = ABC$

Example

Consider the 2^{3-1} design with $C = AB \rightarrow I = ABC$. This is the trivial example because folding over creates the full factorial.

A	B	$C = AB$	Combination	y	1	2	Effect
-	-	+	c	4	15	40	
+	-	-	a	12	24	12	$A + BC$
-	+	-	b	10	8	8	$B + AC$
+	+	+	abc	14	4	-4	$C + AB$

Now we switch signs on everything

A	B	$C = -AB$	Combination	y	1	2	Effect
+	+	-	ab	16	27	41	
-	+	+	bc	11	14	-9	$A - BC$
+	-	+	ac	9	-5	-13	$B - AC$
-	-	-	(1)	5	-4	1	$C - AB$

Obtain all estimates except ABC

Resolution IV Designs

- Can be obtained by folding over resolution III design
- Thus, Resolution IV can be divided into two blocks
- Must contain at least $2k$ runs
- Minimal design if runs equal $2k$
- Determination of defining relation from fold-over III
  - $L + U$ words used as generators
  - L words of like sign
  - U words of unlike sign
  - Combined design with have $L + U - 1$ generators

    All L words
    Even products of U words

Example

- $I = ABD = ACE = BCF = ABCG$ and $I = -ABD = -ACE = -BCF = ABCG$
- $L = 1$ and $U = 3$
- $I = ABCG = ABD(ACE) = ABD(BCF)$
- $I = ABCE = BCDF = ACDG = ABDH = ABCDJ$ and $I = ABCE = BCDF = ACDG = ABDH = -ABCDJ$
- $L = 4$ and $U = 1$
- $I = ABCE = BCDF = ACDG = ABDH$

Problems with Fractional Factorials

- Fractions offer many chances for mistakes
- Too small a design –
  - Tries to estimate too many effects for the number of experimental units used (oversaturation)
- Too large a design –
  - Wasteful of resources
  - Less power

- Aliasing
  - Could wind with misinterpretation of which effects are important
  - Outliers and missing data (order can matter).

3^k Factorial Design

Setting: k factors at 3 levels each (quantitative: low, intermediate, high)

Motivation Find full regression model with quadratic factors

- Not most *efficient* way to model quadratic relationship
- Adding center points and other designs (page 275) might be better

Example: 3^2 Design

2 factors, 3 levels = 9 treatment combinations

- 2 df each for main effects
- 4 df for interaction (fitting linear-linear, linear-quadratic, quadratic-linear, quadratic-quadratic components)

Partitioning Sums of Squares

Partition interaction SS into independent 2-df pieces

- Both pieces estimate interaction component
- Several pieces (as opposed to entire interaction) may be confounded with other effects
- Based on orthogonal LS
- Each piece has no meaning: just separate pieces – all of which estimate interaction effect – which are added together (to get better estimate of interaction, if possible)
- *Interpretation* – if interaction not confounded with main effect (say), is "negligible", then assume piece confounded with main effect is negligible.

Notation

- Partition interaction using superscripts
- *Convention:* first superscript always 1
- **Example**

- Pieces of AB interaction: AB, AB^2
- Pieces of ABC interaction: ABC, ABC^2, AB^2C, and AB^2C^2

- Exponentiate to get other (redundant pieces) which account for second df

$$\begin{aligned}(ABC)^2 &\equiv A^2B^2C^2\\ (ABC^2)^2 &\equiv A^2B^2C\\ (AB^2C)^2 &\equiv A^2BC^2\\ (AB^2C^2)^2 &\equiv A^2BC\end{aligned}$$

Example

- 4 factor experiment – requires 81 runs
- want to reduce to $27 = 3^{4-1}$ runs
- Select (arbitrarily) AB^2CD^2 to confound with identity
- Confounding pattern

$$\begin{array}{rcccl} I &=& AB^2CD^2 &=& A^2BC^2D\\ A &=& (A^2B^2CD^2 = A^4B^4C^2D^4 &=)& ABC^2D\\ &=& (A^3BC^2D &=)& BC^2D\\ B &=& (AB^3CD^2 &=)& ACD^2\\ &=& (A^2B^2C^2D &=)& ABCD^2\\ C &=& AB^2C^2D^2 &=& AB^2D^2\\ D &=& AB^2C &=& AB^2CD\\ AB &=& AC^2D &=& BCD^2\\ AB^2 &=& AB^2C^2D &=& CD^2\\ AC &=& ABCD &=& BD\\ AC^2 &=& ABD &=& BCD\\ AD &=& ABC^2 &=& BC^2D^2\\ AD^2 &=& ABC^2D^2 &=& BC^2\\ BC &=& AC^2D^2 &=& ABD^2\\ BD^2 &=& ACD &=& ABC\\ CD &=& AB^2C^2 &=& AB^2D \end{array}$$

- Resolution IV: 3^{4-1}_{IV}
- **Interpretation:** ABC confounded with D and pieces of AD, BD, and CD interactions.
- Thus, if ABC significant, expect D, AD, BD, and/or CD significant.

ANOVA Table

Source	df
A &/or BCD &/or $ABCD$	2
B &/or ACD &/or $ABCD$	2
C &/or ABD &/or $ABCD$	2
D &/or ABC &/or $ABCD$	2
AB &/or CD &/or ACD &/or BCD &/or $ABCD$	4
AC &/or BD &/or ABD &/or BCD &/or $ABCD$	4
AD &/or BC &/or ABC &/or BCD &/or $ABCD$	4
BC &/or ABD &/or ACD	2
BD &/or ABD &/or ACD	2
CD &/or ABC &/or ACD	2

- Type III SS (might) remove ambiguity
- Resolution IV design – main effects not confounded with 2-factor interactions.

More Than One Identity Generator

Complete generator formed by squaring each individual generator and forming all possible cross-products.

Example

X and Y are generators

$$I = X = X^2 = Y = Y^2 = XY = XY^2 = XY^2 = X^2Y = X^2Y^2$$

- 1/9 fraction: Each term confounded with 8 other terms
- **Example:** Resolution IV Design (generators AB^2CE and AD^2EF)

$$\begin{aligned} I &= AB^2CE = A^2BC^2E^2 = AD^2EF^2 = A^2DE^2F = A^2B^2CD^2E^2F^2 \\ &= B^2CDF = BC^2D^2F^2 = ABC^2DEF \end{aligned}$$

- More complicated examples done by computer

Finding Actual Design Points: Modular Arithmetic

Example

Factors A, B, C, and D (with levels 0, 1, and 2)

If identity generator is $A^\alpha B^\beta C^\gamma D^\delta$, then group contains all values $A^i B^j C^k D^\ell$ such that

$$\alpha i + \beta j + \gamma k + \delta \ell = 0 \mod 3$$

Usually use computer-generated/pre-determined table

Different Blocks

$$\begin{aligned}\alpha i + \beta j + \gamma k + \delta \ell &= 1 \mod 3\\ \alpha i + \beta j + \gamma k + \delta \ell &= 2 \mod 3\end{aligned}$$

General Prime-level Fractions

- Let p be prime number; equal to number of levels each factor contains
- Arithmetic is done $\mod p$.
- Each piece has $p-1$ df
- 2-factor interaction pieces: $AB, AB^2, AB^3, \ldots, AB^{p-1}$
- If X generator, identity generator is

$$I = X = X^2 = \ldots X^{p-1}$$

Mixed Level Designs

- Work with each prime individually and cross resulting designs
- Not desirable, since there are a large number of runs
- May be better to run *Orthogonal Main Effect Design*
  - Not balanced
  - Main effects orthogonal