Session 10

THE 2^K FACTORIAL DESIGN

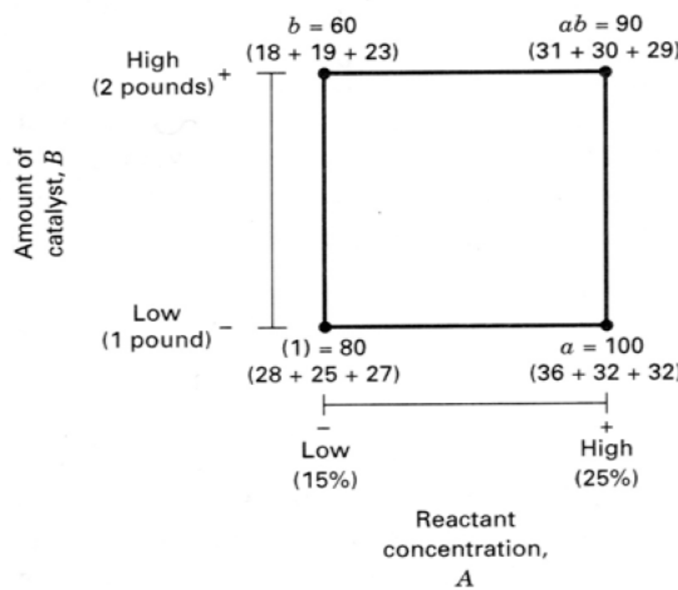
INTRODUCTION

- The 2^k Factorial Design: k factors, each factor has only 2 levels
- Assumptions:
  - The factors are fixed
  - The designs are completely randomized
  - Normality assumptions are satisfied
- The response is assumed linear over the range of the factor levels

THE 2^2 DESIGN

<u>*Example:*</u> Effect of concentration of the reactant (factor A) and the amount of catalyst (factor B) on the yield in a chemical process

A	B	Treatment Combination	Replicate			Total
			I	II	III	
-	-	A low, B low	28	25	27	80
+	-	A high, B low	36	32	32	100
-	+	A low, B high	18	19	23	60
+	+	A high, B high	31	30	29	90



THE 2^2 DESIGN

Notation:

(1), a, b, ab: Total of n replicates at treatment combinations

Main effect

The effect of A at low level of B: $\dfrac{a-(1)}{n}$

The effect of A at high level of B: $\dfrac{ab-b}{n}$

$\Rightarrow$ The main effect of A:

$$\boxed{A = \frac{1}{2n}\left[ab + a - b - (1)\right]}$$

Similarly, the main effect of B:

$$\boxed{B = \frac{1}{2n}\left[ab + b - a - (1)\right]}$$

THE 2^2 DESIGN

Interaction effect

Interaction effect AB is the average difference between the effect of A (or B) at high level of B (or A) and the effect of A (or B) at low level of B (or A).

$$\boxed{AB = \frac{1}{2n}\left[ab + (1) - a - b\right]}$$

Orthogonal Contrasts

$$\boxed{Contrast_A = ab + a - b - (1)}$$

$$\boxed{Contrast_B = ab - a + b - (1)}$$

$$\boxed{Contrast_{AB} = ab - a - b + (1)}$$

THE 2^2 DESIGN

Sums of Squares:

Note that if a contrast C is written in terms of treatment total, $C = \sum_{i=1}^{a} c_i y_{i.}$, then

$$\boxed{SS_C = \frac{C^2}{n\sum_{i=1}^{a} c_i^2}}$$

So:

$$\boxed{SS_A = \frac{\left[ab + a - b - (1)\right]^2}{4n}} \qquad \boxed{SS_B = \frac{\left[ab - a + b - (1)\right]^2}{4n}}$$

$$\boxed{SS_{AB} = \frac{\left[ab - a - b + (1)\right]^2}{4n}}$$

THE 2^2 DESIGN

$$\boxed{SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}} \quad \Rightarrow \quad \boxed{SS_E = SS_T - SS_A - SS_B - SS_{AB}}$$

Example: Effect of concentration of the reactant (factor A) and the amount of catalyst (factor B) on the yield in a chemical process

Contrasts:

$$Contrast_A = (90 + 100 - 60 - 80) = 50$$

$$Contrast_B = (90 - 100 + 60 - 80) = -30$$

$$Contrast_{AB} = (90 - 100 - 60 + 80) = 10$$

The main and interaction effects:

$$A = \frac{Contrast_A}{2n} = 8.33, B = \frac{Contrast_B}{2n} = -5.0, AB = \frac{Contrast_{AB}}{2n} = 1.67$$

THE 2^2 DESIGN

Sums of Squares:

$$SS_A = \frac{(Contrast_A)^2}{4n} = 208.33 \qquad SS_B = \frac{(Contrast_B)^2}{4n} = 75.00$$

$$SS_{AB} = \frac{(Contrast_A)^2}{4n} = 8.33$$

$$SS_T = 323.0 \quad \Rightarrow \quad SS_E = 31.34$$

THE 2^2 DESIGN

Source of Variation	Sum of Squares	DF	Mean Square	F_0	p-value
A	208.33	1	208.33	53.15	0.0001
B	75.00	1	75.00	19.13	0.0024
AB	8.33	1	8.33	2.13	0.1826
Error	31.34	8	3.92		
Total	323.00	11			

Conclusions:

The main effects are significant
There is no interaction between factors.

THE 2^2 DESIGN

Algebraic Notation

Treatment Combination	Factorial Effect I	A	B	AB
(1)	+	-	-	+
a	+	+	-	-
b	+	-	+	-
ab	+	+	+	+

I: represent the *total* or *average* of the entire experiment

THE 2^2 DESIGN

The Regression Model

$$\boxed{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon}$$

$$x_i = \begin{cases} 1 & \text{high level} \\ -1 & \text{low level} \end{cases}$$

Relationship between coded variables and natural variables:

$$x_1 = \frac{\text{Conc} - \left(\text{Conc}_{\text{low}} + \text{Conc}_{\text{high}}\right)/2}{\left(\text{Conc}_{\text{high}} - \text{Conc}_{\text{low}}\right)/2}$$

$$x_2 = \frac{\text{Catalyst} - \left(\text{Catalyst}_{\text{low}} + \text{Catalyst}_{\text{high}}\right)/2}{\left(\text{Catalyst}_{\text{high}} - \text{Catalyst}_{\text{low}}\right)/2}$$

THE 2^2 DESIGN

Least Squares Estimates:

$\hat{\beta}_0$: Grand average of all observations

$\hat{\beta}_1, \hat{\beta}_2$: One-haft of the factor effect estimates

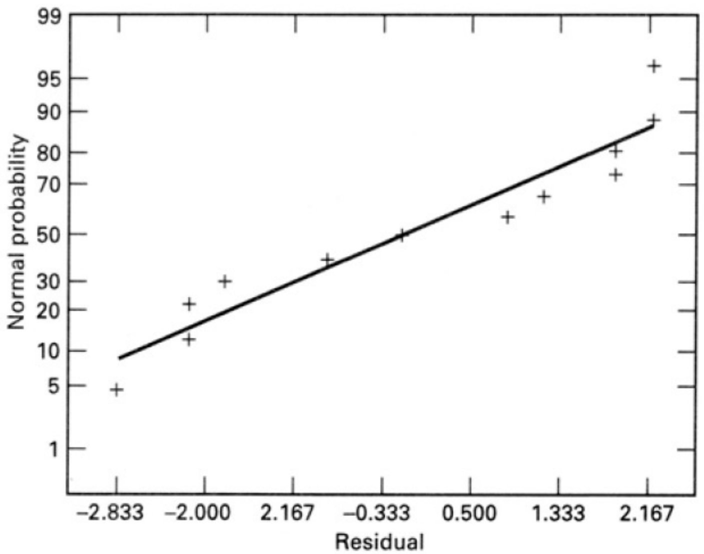
Example: The Chemical Process Experiment

$$\boxed{\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2}$$

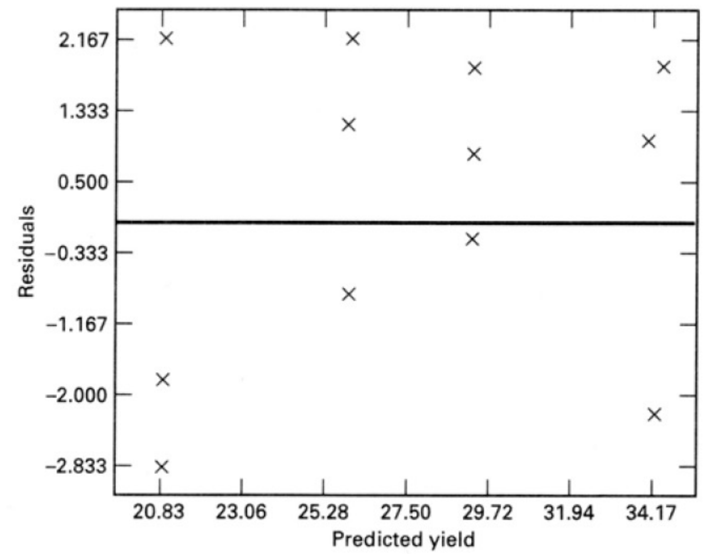
Residuals and Model Adequacy

x_1, x_2	Observation	$\hat{y}$	e
-1,-1	28, 25, 27	25.835	2.165, -0.835, 1.165
1,-1	36, 32, 32	34.165	1.835, -2.165, -2.165
-1,1	18, 19, 23	20.835	-2.835, -1.835, 2.165
1,1	31, 30, 29	29.165	1.835, 0.835, -0.165

THE 2^2 DESIGN



(a) Normal probability plot



(b) Residuals versus predicted yield

No violation of assumptions!

THE 2^2 DESIGN

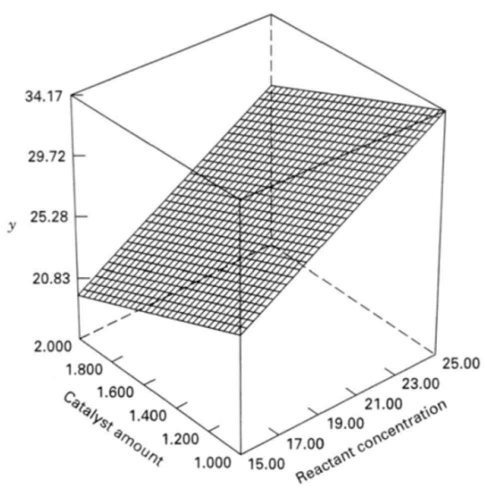
The Response Surface:

The regression model in terms of *natural factor levels*

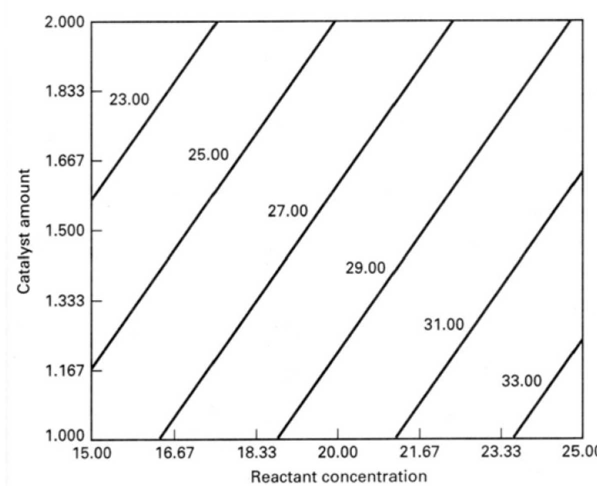
$$\boxed{\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)\left(\frac{\text{Conc} - 20}{5}\right) + \left(\frac{-5.00}{2}\right)\left(\frac{\text{Catalyst} - 1.5}{0.5}\right)}$$

or

$$\hat{y} = 18.33 + 0.8333\text{Conc} - 5.00\text{Catalyst}$$

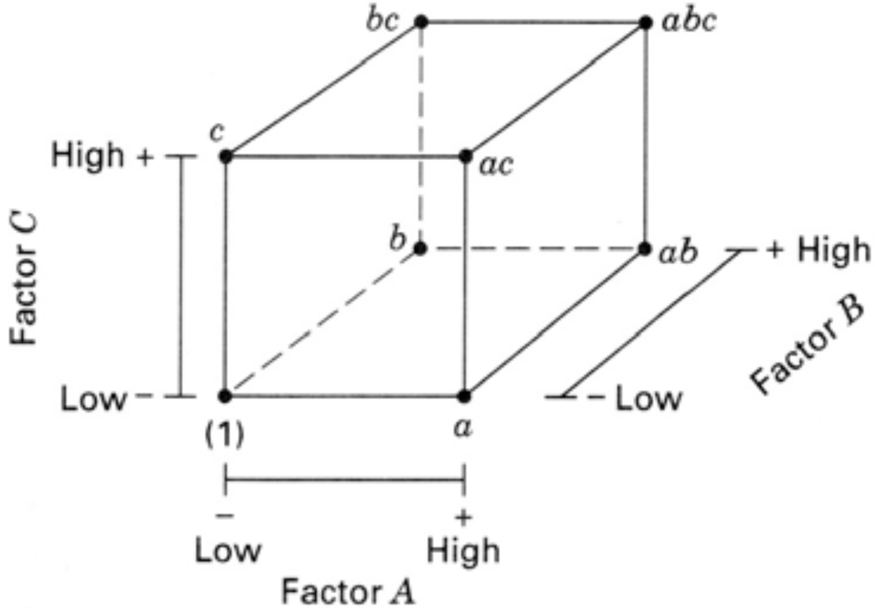


(*a*) Response surface



(*b*) Contour plot

THE 2^3 DESIGN



(a) Geometric view

Run	Factor A	B	C
1	–	–	–
2	+	–	–
3	–	+	–
4	+	+	–
5	–	–	+
6	+	–	+
7	–	+	+
8	+	+	+

(b) The design matrix

THE 2^3 DESIGN

Algebraic Notation

Treatment Combination	Factorial Effect							
	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	-	+	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+	+

THE 2^3 DESIGN

Effect Estimates

$$\boxed{\text{Effect} = \frac{\text{Contrast}}{4n}}$$

For example:

$$\text{Contrast}_{AC} = (1) - a + b - ab - c + ac - bc + abc$$

$$\Rightarrow \qquad AC = \frac{[(1) - a + b - ab - c + ac - bc + abc]}{4n}$$

Sums of Squares

$$\boxed{SS_{Effect} = \frac{(\text{Contrast})^2}{8n}} \quad \left(\sum_{i=1}^{a} c_i^2 = 8\right)$$

$$SS_{AC} = \frac{[(1) - a + b - ab - c + ac - bc + abc]^2}{8n}$$

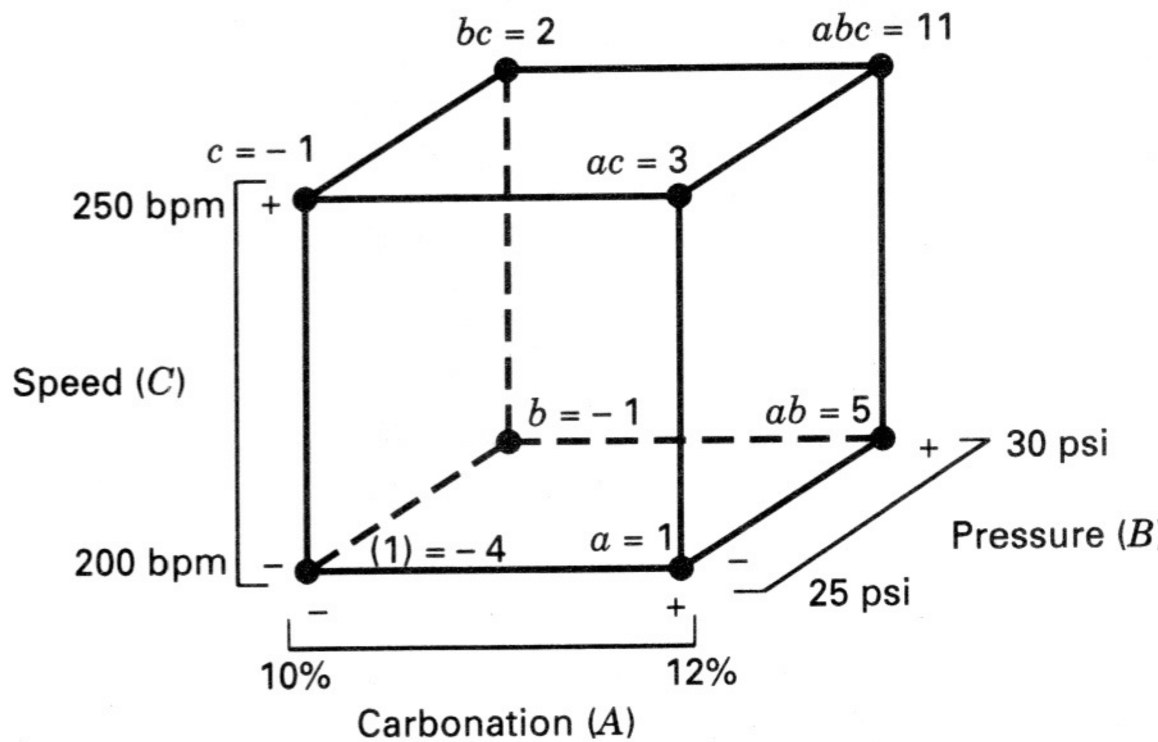
THE 2^3 DESIGN

Example

Investigate the effect on *fill height of a carbonated beverage* of: Percentage of carbonation (Factor *A*); Operating pressure (Factor *B*); Line speed (Factor *C*)

Run	Coded Factors			Fill Height Deviation			Factor Levels	
	A	*B*	*C*	Re.1	Re.2		Low	High
(1)	-	-	-	-3	-1	*A*(%)	10	12
a	+	-	-	0	1	*B*(psi)	25	30
b	-	+	-	-1	0	*C*(b/min)	200	250
ab	+	+	-	2	3			
c	-	-	+	-1	0			
ac	+	-	+	2	1			
bc	-	+	+	1	1			
abc	+	+	+	6	5			

THE 2^3 DESIGN



THE 2^3 DESIGN

Calculation results:

Factor	Effect Estimate	Sum of Squares	Percent Contribution
A	3.00	36.00	46.1538
B	2.25	20.25	25.9615
C	1.75	12.25	15.7051
AB	0.75	2.25	2.8846
AC	0.25	0.25	0.3205
BC	0.50	1.00	1.2821
ABC	0.50	1.00	1.2821
Error		5.00	6.4103
Total		78.00	

THE 2^3 DESIGN

ANOVA

Source of Variation	Sum of Squares	DF	Mean Square	F_0	*p*-value
Perc. Carbonation *A*	36.00	1	36.00	57.60	<0.0001
Pressure *B*	20.25	1	20.25	32.40	0.0005
Line Speed *C*	12.25	1	12.25	19.60	0.0022
AB	2.25	1	2.25	3.60	0.0943
AC	0.25	1	0.25	0.40	0.5447
BC	1.00	1	1.00	1.60	0.2415
ABC	1.00	1	1.00	1.60	0.2415
Error	5.00	8	0.625		
Total	78.00	15			

Conclusions:

Main effects: significant
AB interaction: little significant

THE 2^3 DESIGN

<u>*Regression Model*</u>:

$$\boxed{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_{12} x_1 x_2}$$

$$\Rightarrow \qquad \hat{y} = 1.00 + \left(\frac{3.00}{2}\right)x_1 + \left(\frac{2.25}{2}\right)x_2 + \left(\frac{1.75}{2}\right)x_3 + \left(\frac{0.75}{2}\right)x_1 x_2$$

THE 2^k DESIGN

Structure of Statistical Model

Contain $2^k - 1$ effects, in which:

- k: main effects
- C_k^2: two-factor interaction effects
- C_k^3: three-factor interaction effects
- …
- One k-factor interaction effect

THE 2^k DESIGN

<u>*Standard Order of Treatment Combinations*</u>

Introduce one factor at a time $\Rightarrow$ Standard Order

<u>*Ex*</u>: consider 5 factors; the standard order is:

(1)							
a							
b	ab						
c	ac	bc	abc				
d	ad	bd	abd	cd	acd	bcd	$abcd$
e	ae	be	abe	ce	ace	bce	$abce$
	de	ade	bde	$abde$	cde	$acde$	$bcde$	$abcde$

<u>*Contrast*</u>: $\boxed{\text{Contrast}_{AB\ldots K} = (a \pm 1)(b \pm 1)\ldots(k \pm 1)}$

Use -1 if the factor in included in the effect; use $+1$ otherwise

THE 2^k DESIGN

Ex: Consider the 2^4 design

$$\text{Contrast}_{AC} = (a-1)(b+1)(c-1)(d+1)$$
$$= (1) - a + b - ab - c + ac - bc + abc$$
$$+ d - ad + bd - abd - cd + acd - bcd + abcd$$

Effect Estimates and Sums of Squares

$$\boxed{AB\ldots K = \frac{1}{n2^{k-1}}\left(\text{Contrast}_{AB\ldots K}\right)}$$

$$\boxed{SS_{AB\ldots K} = \frac{1}{n2^{k}}\left(\text{Contrast}_{AB\ldots K}\right)^2}$$

A SINGLE REPLICATE 2^k DESIGN

Why unreplicated factorial design?

- Total number of treatment combinations is large
- Resources are limited

Problem

No internal estimate of pure error!

$\Rightarrow$ Assume that some high-order interactions are negligible and combine their mean squares to estimate error

A SINGLE REPLICATE 2^k DESIGN

The Sparsity of Effects Principle:

Most systems are dominated by some of the main effects and low-order interactions, most high-order interaction are negligible.

Daniel's Approach

Examine the *normal probability plot* of the effect estimates: the effects that are negligible are normally distributed and hence, tend to fall along a straight line.

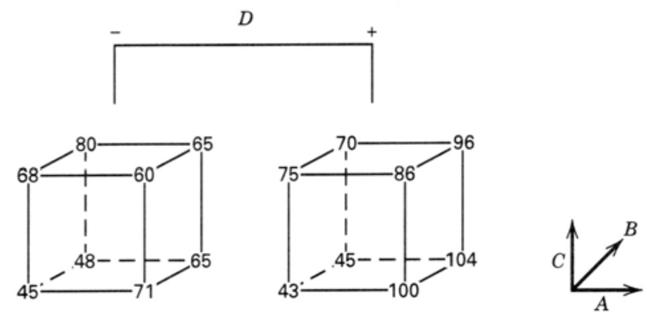
Example: Filtration Rate Experiment

Consider the production of a chemical product in a pressure vessel. Four factors that may have effects on the filtration rate are:

- Temperature (A)
- Pressure (B)
- Concentration of formaldehyde (C)
- Stirring rate (D)

A SINGLE REPLICATE 2^k DESIGN

Run No.	Factor				Run Label	Filtration Rate (gal/h)
	A	*B*	*C*	*D*		
1	-	-	-	-	(1)	45
2	+	-	-	-	*a*	71
3	-	+	-	-	*b*	48
4	+	+	-	-	*ab*	65
5	-	-	+	-	*c*	68
6	+	-	+	-	*ac*	60
7	-	+	+	-	*bc*	80
8	+	+	+	-	*abc*	65
9	-	-	-	+	*d*	43
10	+	-	-	+	*ad*	100
11	-	+	-	+	*bd*	45
12	+	+	-	+	*abd*	104
13	-	-	+	+	*cd*	75
14	+	-	+	+	*acd*	86
15	-	+	+	+	*bcd*	70
16	+	+	+	+	*abcd*	96

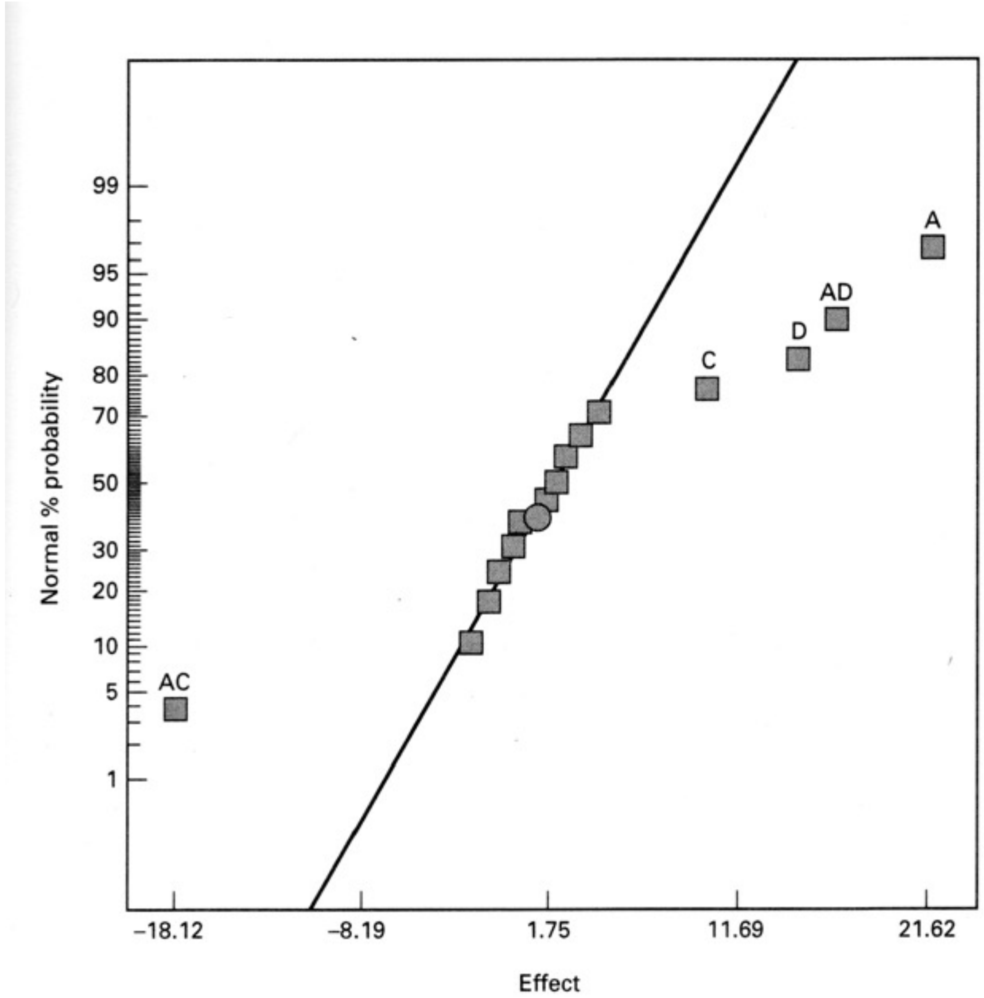


A SINGLE REPLICATE 2^k DESIGN

Effect estimates and sums of squares:

Effect	Effect Estimate	SS	Perc. Contribution
A	21.625	1870.56	32.6397
B	3.125	39.0625	0.681608
C	9.875	390.062	6.80626
D	14.625	855.563	14.9288
AB	0.125	0.0625	0.00109057
AC	-18.125	1314.06	22.9293
AD	16.625	1105.56	19.2911
BC	2.375	22.5625	0.393696
BD	-0.375	0.5625	0.00981515
CD	-1.125	5.0625	0.0883363
ABC	1.875	14.0625	0.245379
ABD	4.125	68.0625	1.18763
ACD	-1.625	10.5625	0.184307
BCD	-2.625	27.5625	0.480942
$ABCD$	1.375	7.5625	0.131959

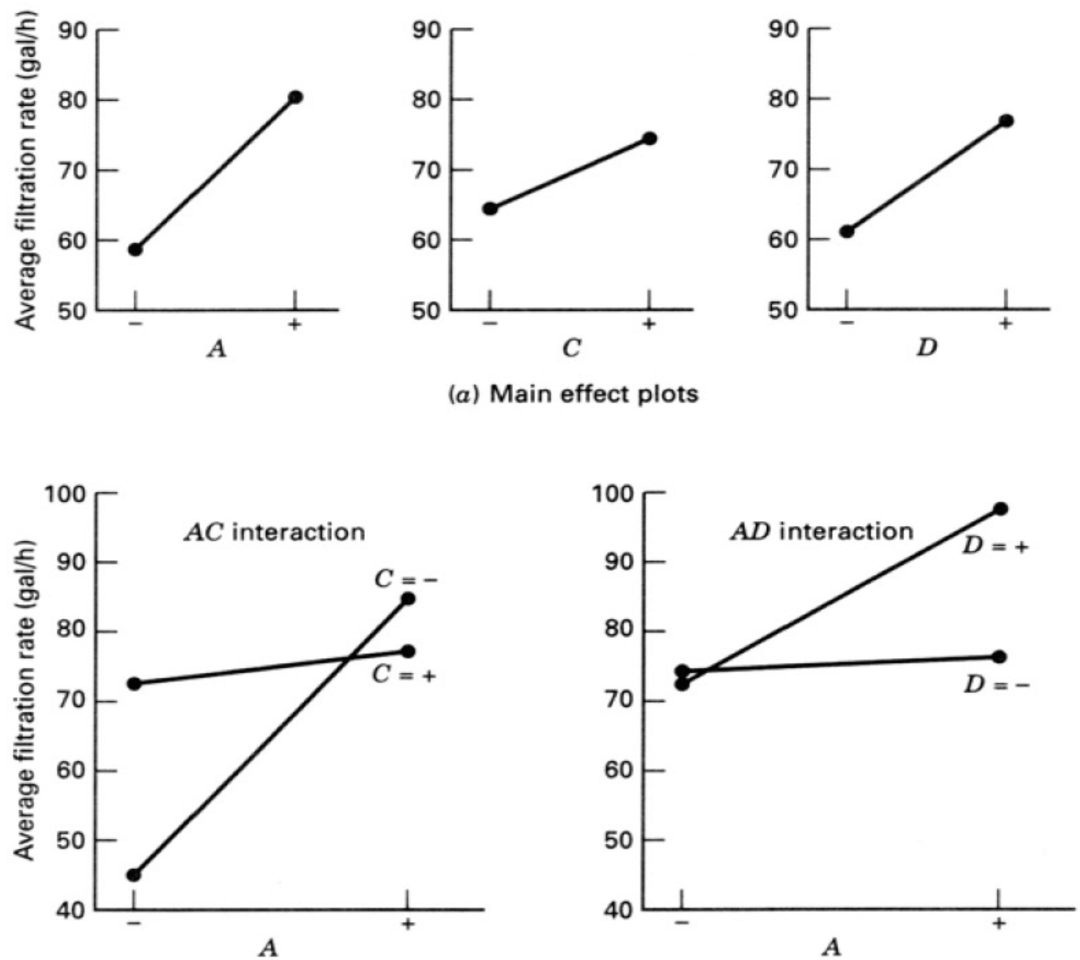
A SINGLE REPLICATE 2^k DESIGN



Normal Probability Plot of Effect Estimates

A SINGLE REPLICATE 2^k DESIGN

The important effects are: A, C, D, AC, AD



(a) Main effect plots

(b) Interaction plots

A SINGLE REPLICATE 2^k DESIGN

Notes:

- All main effects are positive $\Rightarrow$ We would run all these factors at high levels !!?

- From AC interaction:
  - Temperature effect is very small when concentration is at high level
  - Temperature effect is very large when concentration is at low level
  - Best result obtained at low concentration and high temperature

- From AD interaction:
  - Stirring rate has little effect at low temperature and large positive effect at high temperature.

So, the best decision might be: A&D – high level, C – low level.

A SINGLE REPLICATE 2^k DESIGN

ANOVA

Source of Variation	Sum of Squares	DF	Mean Square	F_0	p-value
A	1870.56	1	1870.56	83.36	<0.0001
C	390.06	1	390.06	17.38	<0.0001
D	855.56	1	855.56	38.13	<0.0001
AC	1314.06	1	1314.06	58.56	<0.0001
AD	1105.56	1	1105.56	49.27	<0.0001
CD	5.06	1	5.06	<1	
ACD	10.56	1	10.56	<1	
Error	179.52	8	22.44		
Total	5730.94	15			

Regression Model:

$$\hat{y} = 70.06 + \left(\frac{21.625}{2}\right)_1 x_1 + \left(\frac{9.875}{2}\right) x_3 + \left(\frac{14.625}{2}\right) x_4 - \left(\frac{18.125}{2}\right) x_1 x_3 + \left(\frac{16.625}{2}\right) x_1 x_4$$

(Note: 70.06 = average response)

ADDITION OF CENTER POINT IN 2^k DESIGN

Consider the assumption of *linearity* in the factor effects:

❖ If interaction terms are included in the model

$$\boxed{y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j} \beta_{ij} x_i x_j + \varepsilon}$$

The model is *capable* of representing some curvature in the response function

$\Rightarrow$ No need *perfect* linearity

ADDITION OF CENTER POINT IN 2^k DESIGN

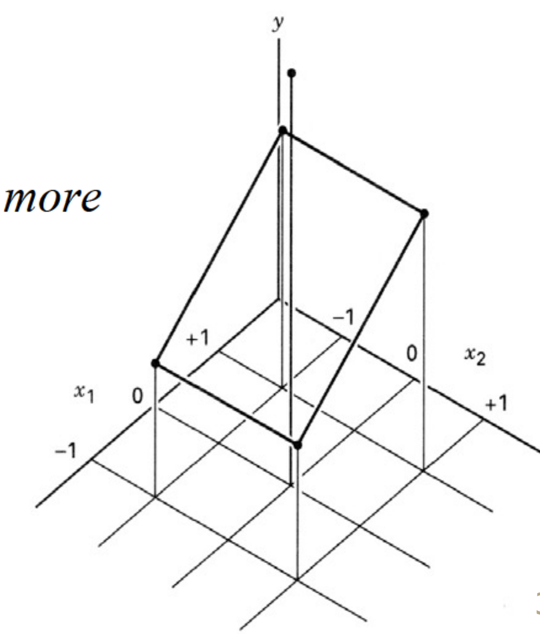
❖ In case the above model cannot be applied, the ***second-order response surface model*** with quadratic effects might be an alternative:

$$\boxed{y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \varepsilon}$$

Why use additional center point?

To detect if the second-order response surface model is more appropriate and

To obtain an independent estimate of error



ADDITION OF CENTER POINT IN 2^k DESIGN

Testing Curvature Effects

- Compute sum of squares for quadratic curvature:

$$\boxed{SS_{\text{Pure Quadratic}} = \frac{n_F n_C \left(\overline{y}_F - \overline{y}_C\right)^2}{n_F + n_C}} \quad (d.f.=1)$$

in which:

n_F : number of factorial design points

n_C : number of center points

- Compute error estimate based on center points:

$$\boxed{MS_E = \frac{SS_E}{df_E} = \frac{\sum_{\text{Center Points}} \left(y_i - \overline{y}_C\right)^2}{n_C - 1}}$$

- Compare $MS_{\text{Pure Quadratic}}$ with MS_E to test for quadratic curvature

ADDITION OF CENTER POINT IN 2^k DESIGN

Example: A chemical engineer is studying the yield of a process.
Factors of interest: Reaction time (A) – Reaction temperature (B)

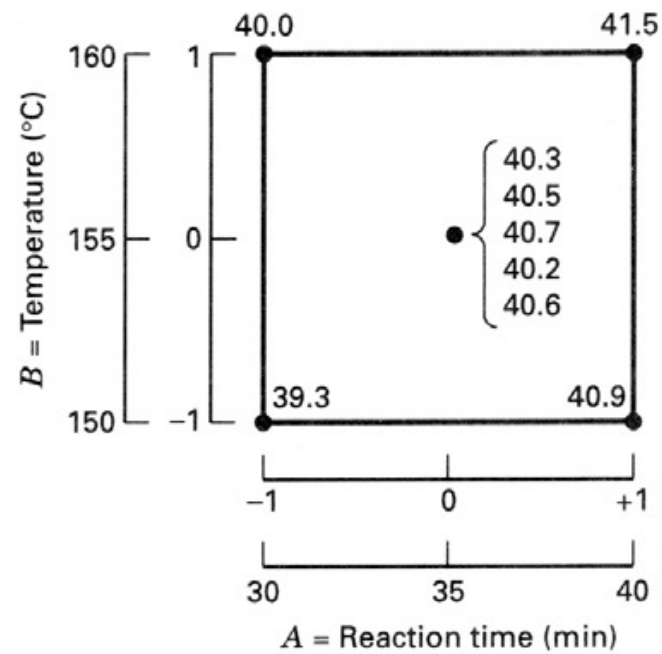
Due to uncertainty about linearity over the region of exploration, an unreplicated 2^2 design with 5 center points is conducted.

$$SS_{\text{Pure Quadratic}} = \frac{n_F n_C \left(\bar{y}_F - \bar{y}_C\right)^2}{n_F + n_C}$$

$$= \frac{(4)(5)(40.425 - 40.46)^2}{4+5}$$

$$= 0.0027$$

$$MS_E = \frac{\sum_{i=1}^{5} (y_i - 40.46)^2}{5-1} = 0.043$$



ADDITION OF CENTER POINT IN 2^k DESIGN

Source of Variation	Sum of Squares	DF	Mean Square	F_0	p-value
A (Time)	2.4025	1	2.4025	55.87	0.0017
B (Temperature)	0.4225	1	0.4225	9.83	0.0350
AB	0.0025	1	0.0025	0.06	0.8185
Pure Quadratic	0.0027	1	0.0027	0.06	0.8185
Error	0.1720	4	0.0430		
Total	3.0022	8			

Conclusions:

- Main effects are significant - There is no interaction
- No evidence of second-order curvature in the response, i.e., the null hypothesis $H_0 : \beta_{11} + \beta_{22} = 0$ cannot be rejected.

CONFOUNDING IN THE 2^k DESIGN

Why confounding? *It is impossible to perform a complete replicate of a factorial design in one block*

What is confounding? *A design technique for arranging a complete factorial experiment in blocks*

- ❖ The block size is smaller than the number of treatment combinations in one replicate
- ❖ Information about some *high-order interactions* is *indistinguishable from blocks* (*confounded with blocks*).

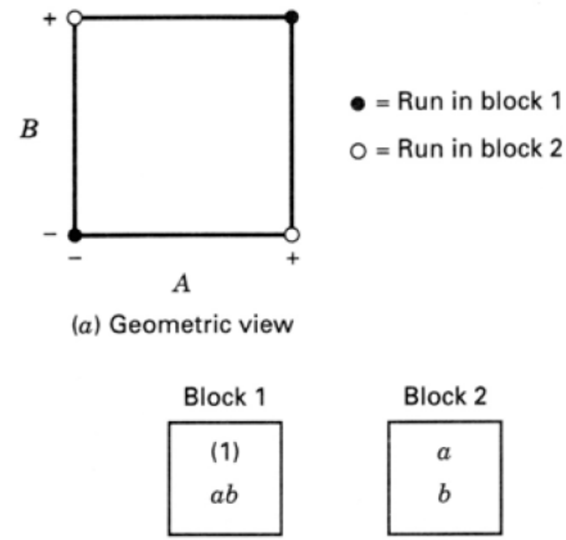
Usually, the technique is use for 2^k design in 2^p blocks.

CONFOUNDING IN TWO BLOCKS

Consider a 2^2 design in which each batch of material (blocks) is only large enough for two treatment combinations.

Arrange the two blocks as follows:

Block 1: (1) and ab
Block 2: a and b



(a) Geometric view

(b) Assignment of the four runs to two blocks

CONFOUNDING IN TWO BLOCKS

Consider effect estimates:

$$A = \frac{1}{2}\left[ab + a - b - (1)\right]$$

$$B = \frac{1}{2}\left[ab - a + b - (1)\right]$$

$$AB = \frac{1}{2}\left[ab - a - b + (1)\right]$$

Note that

- A, B are unaffected by blocking: in each estimate there is one + and one – treatment combination from each block
- AB is confounded with blocks: the block effect and the AB interaction are identical.

CONFOUNDING IN TWO BLOCKS

Procedure for constructing the block

In order to confound an effect with block:

Assign the treatment combinations with *plus* sign to block 1
Assign the treatment combinations with *minus* sign to block 2

For example: To confound main effect A with block

Assign (1) and b to block 1
Assign a and ab to block 2

CONFOUNDING IN TWO BLOCKS

Other methods for constructing the blocks

1. Use the ***Defining Contrast***:

$$\boxed{L = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k}$$

$x_i = 0,1$: level of the ith factor appearing in a treatment combination
$\alpha_i = 0,1$: the exponent appearing on the ith factor in the effect to be confounded

Treatment combination that produce the same value of $L(\text{mod } 2)$ will be assigned in the same block.

CONFOUNDING IN TWO BLOCKS

Other methods for constructing the blocks

1. Use the ***Defining Contrast***:

Example: consider 2^3 design with ABC confounded with blocks

Treatment Combination	Factorial Effect							
	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	-	+	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	+	-	+	-	-
bc	+	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+	+

CONFOUNDING IN TWO BLOCKS

Other methods for constructing the blocks

1. Use the ***Defining Contrast***:

We have: $x_1 \to A, x_2 \to B, x_3 \to C$ $\alpha_1 = \alpha_2 = \alpha_3 = 1$ Hence $L = x_1 + x_2 + x_3$

Value of L:

(1):	$L = 0+0+0 = 0 \pmod 2$	a:	$L = 1+0+0 = 1 \pmod 2$
ab:	$L = 1+1+0 = 0 \pmod 2$	b:	$L = 0+1+0 = 1 \pmod 2$
ac:	$L = 1+0+1 = 0 \pmod 2$	c:	$L = 0+0+1 = 1 \pmod 2$
bc:	$L = 0+1+1 = 0 \pmod 2$	abc:	$L = 1+1+1 = 1 \pmod 2$
	BLOCK 1		**BLOCK 2**

CONFOUNDING IN TWO BLOCKS

Other methods for constructing the blocks

2. **Principle Block**

- The principle block is the block contain (1)
- Any other element in the principle block may be generated by multiplying two other elements in the principle block *modulus* 2.
- Treatment combinations in other blocks may be generated by *multiplying* one element in the new block by each element in the principle block *modulus* 2

Example: consider 2^3 design with ABC confounded with blocks

Block 2 – principle block – contains (1), ab, ac, bc due to:

$$ab.ac = a^2bc = bc \qquad ab.bc = ab^2c = ac \qquad ac.bc = abc^2 = ab$$

Block 2 should contain a, therefore it includes

$$ab.a = a^2b = b \quad ac.a = a^2c = c \quad bc.a = abc$$

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

- When the number of factor is small (k = 2, 3), it is necessary to *replicate* the experiment to obtain an estimate of error.

Consider a 2^3 factorial design with 4 replicates

Replication 1		Replication 2		Replication 3		Replication 4	
B1	B2	B1	B2	B1	B2	B1	B2
(1)=-3	abc=6	(1)=-1	abc=5	(1)=-2	abc=4	(1)=-3	abc=4
ac=2	a=0	ac=1	a=1	ac=2	a=1	ac=1	a=0
ab=2	b=-1	ab=3	b=0	ab=2	b=-1	ab=3	b=0
bc=1	c=-1	bc=1	c=0	bc=2	c=1	bc=2	c=1

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

The source of variation and degrees of freedom are:

Source of Variation	DF
Replicates	3
Block (*ABC*)	1
Error for *ABC* (replicates$\times$blocks)	3
A	1
B	1
C	1
AB	1
AC	1
BC	1
Error (replicates$\times$Effects)	18
Total	31

ANOVA: The students are required to work on this problem!

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

- When the number of factor is large (k = 4, 5, …), a single *replicate* can be applied and error can be estimated from the sums of squares of high-order interactions.

Example: Reconsider the Filtration Rate Experiment (slide 27)

Four factors that have effects on the filtration rate are:

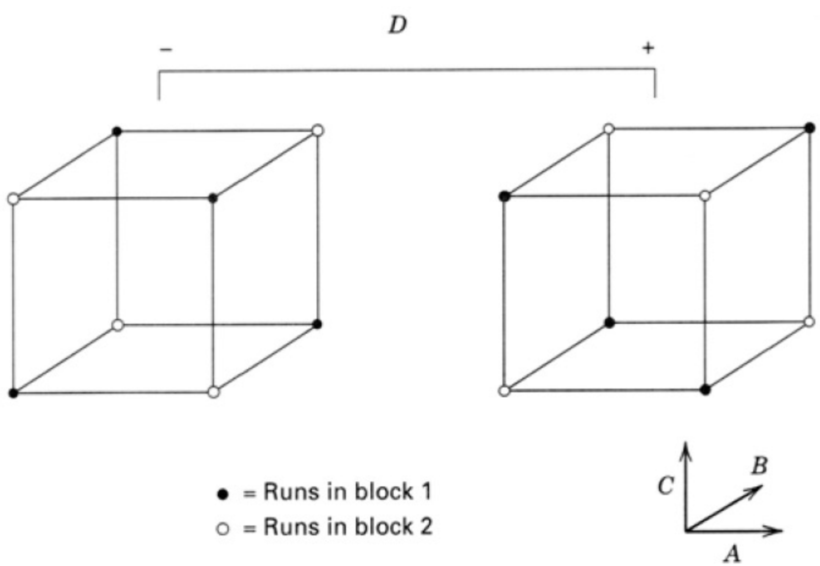
- Temperature (A)
- Pressure (B)
- Concentration of formaldehyde (C)
- Stirring rate (D)

Suppose that only 8 treatment combinations can be run from a single batch of material: 2 batches are required ➔ 2 blocks

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

Suppose also that batch 1 is of poor quality and all responses will be 20 units lower in this batch than in the other



(*a*) Geometric view

Block 1	Block 2
(1) = 25	a = 71
ab = 45	b = 48
ac = 40	c = 68
bc = 60	d = 43
ad = 80	abc = 65
bd = 25	bcd = 70
cd = 55	acd = 86
$abcd$ = 76	abd = 104

(*b*) Assignment of the 16 runs to two blocks

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

Effect Estimates

Effect	Effect Estimate	SS	Perc. Contribution
A	21.625	1870.56	26.30
B	3.125	39.0625	0.55
C	9.875	390.062	5.49
D	14.625	855.563	12.03
AB	0.125	0.0625	<0.01
AC	-18.125	1314.06	18.48
AD	16.625	1105.56	15.55
BC	2.375	22.5625	0.32
BD	-0.375	0.5625	<0.01
CD	-1.125	5.0625	0.07
ABC	1.875	14.0625	0.20
ABD	4.125	68.0625	0.96
ACD	-1.625	10.5625	0.15
BCD	-2.625	27.5625	0.39
Blocks (*ABCD*)	**-18.625**	**1387.5625**	**19.51**

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

Note that the only change here is in *ABCD* effects. It is reduced by 20 from the original value (1.375). This effect is actually the estimate of (Block + *ABCD*) and can be calculated directly from

$$\begin{aligned}\text{Block Effect} &= \bar{y}_{Block1} - \bar{y}_{Block2} \\ &= \frac{406}{8} - \frac{555}{8} = -18.625\end{aligned}$$

(Note: We can apply contrast of *ABCD*!)

The block sum of squares:

$$SS_{Blocks} = \frac{406^2 + 555^2}{8} - \frac{(406 + 555)^2}{16} = 1387.5625$$

(Note: We can apply contrast of *ABCD*!)

CONFOUNDING IN TWO BLOCKS

Estimation of Error:

ANOVA

Source of Variation	Sum of Squares	DF	Mean Square	F_0	p-value
Blocks ($ABCD$)	1387.5625	1			
A	1870.5625	1	1870.56	89.76	<0.0001
C	390.0625	1	390.06	18.72	0.0019
D	855.5625	1	855.56	41.05	0.0001
AC	1314.0625	1	1314.06	63.05	<0.0001
AD	1105.5625	1	1105.56	53.05	<0.0001
Error	187.5625	9	20.84		
Total	7111.4375	15			

Notes:

1. SS_E : Total sums of squares of B, AB, BC, BD, CD, ABC, ABD, ACD, and BCD.
2. The conclusions exactly match former conclusions when no block effect was present.

CONFOUNDING IN FOUR BLOCKS

Consider the 2^5 design confounded in 4 blocks. The effects to be confounded with block are ADE and BCE.

Defining Contrasts:

$$L_1 = x_1 + x_4 + x_5$$
$$L_2 = x_2 + x_3 + x_5$$

The four blocks are:

Block 1 $L_1 = 0$ $L_2 = 0$		Block 2 $L_1 = 1$ $L_2 = 0$		Block 3 $L_1 = 0$ $L_2 = 1$		Block 4 $L_1 = 1$ $L_2 = 1$	
(1)	abc	a	be	b	$abce$	e	$abcde$
ad	ace	d	$abde$	abd	ae	ade	bd
bc	cde	abc	ce	c	$bcde$	bce	ac
$abcd$	bde	bcd	$acde$	acd	de	ab	cd

CONFOUNDING IN FOUR BLOCKS

Note that

- We have 4 block with d.f. = 3
- Each *ADE* & *BCE* has d.f. = 1

$\Rightarrow$ There should be another effect confounded with block: the *generalized interaction* of *ADE* and *BCE*.

$$(ADE)(BCE) = ABCDE^2 = ABCD \pmod 2$$

In selecting the effects to be confounded with blocks, effects that may be of interest should not be confounded with blocks.

For example, if the two effects selected to be confounded with blocks are *ABCDE* and *ABD* then the other confounded effect will be *CE*. This design is not good because *CE* might be an effect of interest.

CONFOUNDING IN 2^P BLOCKS

Procedure for constructing blocks:

1. Select p *independent* effects to be confounded – "independent": no effect chosen is the generalized interaction of the others.

2. The blocks are generated by use of p defining contrasts

3. Find $\left(2^p - p - 1\right)$ other confounded effects

Example: Consider the 2^5 design confounded in 8 blocks ($p = 3$)

1. Select 3 confounded effects: *ABEF*, *ABCD*, *ACE*
2. Generate the blocks
3. Find other $\left(2^p - p - 1\right) = 4$ confounded effects:

$(ABEF)(ABCD) = CDEF$ $\quad$ $(ABEF)(ACE) = BCF$

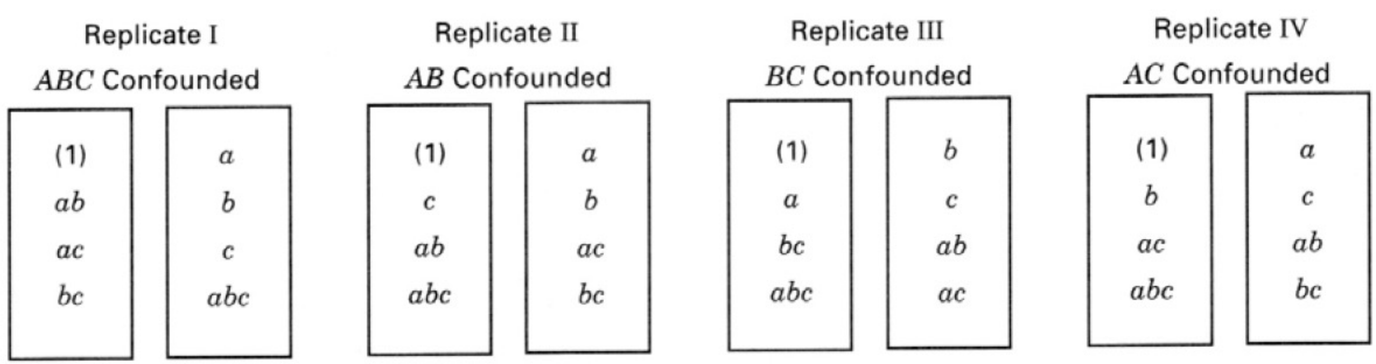
$(ABCD)(ACE) = BDE$ $\quad$ $(ABEF)(ABCD)(ACE) = ADF$

PARTIAL CONFOUNDING

- In former sessions, we dealt with *Completely Confounded Design* and hence, information from confounded effects cannot be retrieved.

- In cases when resources are not so limited and replication can be used in the design to estimate error, it is also better to confound *difference effects in different replicates* $\Rightarrow$ Partial Confounding

<u>*Example*</u>: 2^3 design with 4 replicates and different effects are confounded in each replicate.



Partial confounding in the 2^3 design.

PARTIAL CONFOUNDING

With this design:

- Information on ABC can be obtained from Rep.II, Rep.III, Rep.IV
- Information on AB can be obtained from Rep.I, Rep.III, Rep.IV
- Information on BC can be obtained from Rep.II, Rep.II, Rep.IV
- Information on AC can be obtained from Rep.II, Rep.II, Rep.III

PARTIAL CONFOUNDING

Degrees of Freedom and Sums of Squares:

- The degrees of freedom are presented in the below table
- The sums of squares can be computed from corresponding contrasts, in which the error sum of squares is the sum of

  1. Sums of squares of $replicates \times main\ effects$
  2. Sums of squares of $replicates \times interaction\ effects$ for each replicate in which the interaction effect is unconfounded

  For example, $replicates \times ABC$ for replicates II, III and IV

PARTIAL CONFOUNDING

Source of Variation	DF
Replicates	3
Blocks within Replicates	
ABC (rep.1) + AB (rep.2) + BC (rep.3) + AC (rep.4)	4
A	1
B	1
C	1
AB (from replicates I, III and IV)	1
AC (from replicates I, II and III)	1
BC (from replicates I, II and IV)	1
ABC (from replicates II, III and IV)	1
Error	17
Total	31

PARTIAL CONFOUNDING

Example: Consider the fill height of a carbonated beverage experiment

Factors: Percentage of carbonation (A)
Operating pressure (B)
Line speed (C)

The data for a partial confounded design in two blocks are:

Replicate I
ABC confounded

Block 1	Block 2
(1) = -3	$a = 0$
$ab = 2$	$b = -1$
$ac = 2$	$c = -1$
$bc = 1$	$abc = 6$

$R_1 = 6$

Replicate II
AB confounded

Block 1	Block 2
(1) = -1	$a = 1$
$c = 0$	$b = 0$
$ab = 3$	$ac = 1$
$abc = 5$	$bc = 1$

$R_2 = 10$

PARTIAL CONFOUNDING

- The sums of squares for A, B, C, AC, and BC are calculated as usual (use contrasts)
- SS_{ABC} is calculated using only data from replicate II

$$SS_{ABC} = \frac{\left[-(1)+a+b-ab+c-ac-bc+abc\right]^2}{n2^k} =$$

$$= \frac{\left[-(-1)+1+0-3+0-1-1+5\right]^2}{(1)(8)} = 0.50$$

- SS_{AB} is calculated using only data from replicate I

$$SS_{AB} = \frac{\left[(1)-a-b+ab+c-ac-bc+abc\right]^2}{n2^k} =$$

$$= \frac{\left[(-3)-0-(-1)+2+(-1)-2-1+6\right]^2}{(1)(8)} = 0.50$$

PARTIAL CONFOUNDING

- SS_{Rep} is calculated as follows:

$$SS_{\text{Rep}} = \sum_{h=1}^{n} \frac{R_h^2}{2^k} - \frac{y_{\dots}^2}{N} = \frac{6^2 + 10^2}{8} - \frac{16^2}{16} = 1.00$$

- Sum of squares of blocks within replicated is the sum of SS_{ABC} from replicate I and SS_{AB} from replicate II, in which

$$SS_{ABC(\text{I})} = \frac{\left[-(1) + a + b - ab + c - ac - bc + abc\right]^2}{n2^k} =$$

$$= \frac{\left[-(-3) + 0 + (-1) - 2 + (-1) - 2 - 1 + 6\right]^2}{(1)(8)} = 0.50$$

$$SS_{AB(\text{II})} = \frac{\left[(1) - a - b + ab + c - ac - bc + abc\right]^2}{n2^k} =$$

$$= \frac{\left[(-1) - 1 - 0 + 3 + 0 - 1 - 1 + 5\right]^2}{(1)(8)} = 2.00$$

PARTIAL CONFOUNDING

ANOVA

Source of Variation	Sum of Squares	DF	Mean Square	F_0	p-value
Replicates	1.00	1	1.00	-	
Blocks within replicates	2.50	2	1.25	-	
A	36.00	1	36.00	48.00	0.0001
B	20.25	1	20.25	27.00	0.0035
C	12.25	1	12.25	16.33	0.0099
AB (rep.I only)	0.50	1	0.50	0.67	0.4503
AC	0.25	1	0.25	0.33	0.5905
BC	1.00	1	1.00	1.33	0.3009
ABC (rep.II only)	0.50	1	0.50	0.67	0.4503
Error	3.75	5	0.75	-	
Total	78.00	15		-	

Conclusions: Only main effects are significant!

PARTIAL CONFOUNDING

Assignment:

Let compute the error sum of square in the above example directly

Hint: interaction effect can be compute from the following formula

$$\boxed{n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)^2}$$