01NAEX - Lecture 07
2^k Factorial Design (2)
Center point, Blocking, Confounding

Jiri Franc

Czech Technical University
Faculty of Nuclear Sciences and Physical Engineering
Department of Mathematics

Last Lesson - A single Replicate 2^k Designs

- 2^k factorial design is a design with one observation at each corner of the ”cube” and sometimes is called single replicated design.
- Very widely used type of design, especially in first planning and testing.
- If the factors are spaced too closely, it increases the chances that the noise will overwhelm the signal in the data.
- More aggressive spacing is usually best.
- Lack of replication causes potential problems in statistical testing:
  - With no replication, fitting the full model results in zero degrees of freedom for error.
  - Omit higher-order interactions and estimate error.
  - Normal probability plotting of effects (Daniels,1959).
  - Lenth’s method - Pareto plot (Lenth, 1989).
- Create contour plots and find optimal settings for numerical variables.

2^k Designs are Optimal Designs

Two-level factorial designs have many interesting properties. One of the most useful is that 2^k designs minimize the variance of the model regression coefficients.

Consider a simple case the 2^2 design with one replicate and coded variables.

- The regression model fitted to the data is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon.$$

- Four runs in this design in terms of the regression model are

$$\begin{aligned}
(1) &= \beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_{12}(-1)(-1) + \varepsilon_1 \\
a &= \beta_0 + \beta_1(1) + \beta_2(-1) + \beta_{12}(1)(-1) + \varepsilon_2 \\
b &= \beta_0 + \beta_1(-1) + \beta_2(1) + \beta_{12}(-1)(1) + \varepsilon_3 \\
ab &= \beta_0 + \beta_1(1) + \beta_2(1) + \beta_{12}(1)(1) + \varepsilon_4
\end{aligned}$$

- In matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

2^k Designs are Optimal Designs (2)

Regression coefficients β's are estimated by the Ordinary Least Squares (OLS)

$$\hat{\beta}^{(OLS)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

where

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} (1) \\ a \\ b \\ ab \end{bmatrix} = \begin{bmatrix} (1) + a + b + ab \\ -(1) + a - b + ab \\ -(1) - a + b + ab \\ (1) - a - b + ab \end{bmatrix}$$

The $\mathbf{X}'\mathbf{X}$ matrix is diagonal because the 2^k designs are orthogonal.

2^k Designs are Optimal Designs (3)

In 2^2 example the regression coefficients β's are

$$\hat{\boldsymbol{\beta}}^{(OLS)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} \frac{(1)+a+b+ab}{4} \\ \frac{-(1)+a-b+ab}{4} \\ \frac{-(1)-a+b+ab}{4} \\ \frac{(1)-a-b+ab}{4} \end{pmatrix}$$

The least squares estimates of the model regression coefficients are exactly equal to one-half of the usual effect estimates!

It turns out that the variance of any model regression coefficient is easy to find

$$Var(\hat{\beta}) = \frac{\sigma^2}{\text{diagonal elements of } \mathbf{X}'\mathbf{X}} = \frac{\sigma^2}{n2^k} = \frac{\sigma^2}{N}.$$

All model regression coefficients have the same variance and this is the minimum possible variance for the regression coefficient.

2^k Designs are Optimal Designs (3)

For fitting the first-order model or the first-order model with interaction, the 2^k design is (equations for 2^2 design)

Linear model: $y = X\beta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 I)$. Information: $M = X^\top X$.

- D-optimal:
  - design minimizes the variance of the model regression coefficients.
  - maximize $\det(M) \Rightarrow$ minimize volume of the joint CI ellipsoid for β. Efficient estimation
  - Adding center helps curvature check but may slightly reduce $\det(M)$ for pure first-order fitting.

2^k Designs are Optimal Designs (3)

For fitting the first-order model or the first-order model with interaction, the 2^k design is (equations for 2^2 design)

Linear model: $y = X\beta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 I)$. Information: $M = X^\top X$.

- G-optimal:
  - the model that we fit to the data from the experiment minimizes the maximum prediction variance over the design region.

$$
\begin{aligned}
Var(\hat{y}(x_1, x_2)) &= Var(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2) \\
&= \frac{\sigma^2}{4}(1 + x_1 + x_2 + x_1 x_2)
\end{aligned}
$$

  - minimize $\max_{x \in \mathcal{X}} \mathrm{Var}\{\hat{y}(x)\} = \sigma^2\, x^\top M^{-1} x$. Controls the *worst* prediction variance.

2^k Designs are Optimal Designs (3)

For fitting the first-order model or the first-order model with interaction, the 2^k design is (equations for 2^2 design)

Linear model: $y = X\beta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 I)$. Information: $M = X^\top X$.

- I-optimal:
  - the smallest possible value of the average prediction variance that can be obtained from a k-run design.

$$\begin{aligned} I &= \frac{1}{A}\int_{-1}^{1}\int_{-1}^{1} Var(\hat{y}(x_1, x_2))dx_1 dx_2 \\ &= \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \frac{\sigma^2}{4}(1 + x_1 + x_2 + x_1 x_2)dx_1 dx_2 = \frac{4\sigma^2}{9}, \end{aligned}$$

    where A is the volume of the design space.
  - minimize $\int_{\mathcal{X}} \mathrm{Var}\{\hat{y}(x)\}\, d\mu(x)$. Minimizes *average* prediction variance over region $\mathcal{X}$.

Heuristic: D for precise β, I/G for precise $\hat{y}$ (I = average, G = worst-case). Choose region $\mathcal{X}$ that matches your process window.

Addition of Center Points to a 2^k Designs

In two level factorial design we assume linearity in the factor effects.
If the interaction is presented in the model, the curvature in the response function can be modeled by different ways.

First-order model (interaction)

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

Second-order response surface model

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

Addition of Center Points

- Based on the idea of replicating some of the runs in a factorial design.
- Runs at the center provide an estimate of error and allow the experimenter to distinguish between models.

Addition of Center Points to a 2^k Designs

The method is based on replicating certain point in a 2^k factorial that will provide protection against curvature from the second order effects as well as allow an independent estimate of error to be obtained.

- Center point consists of n_C replicates run in $x_1 = x_2 = \ldots = x_k = 0$.
- k factors have to be **quantitative**.
- Center points do not affects the usual estimates in 2^k design.

Addition of Center Points to a 2^k Designs

Compute

- $\bar{y}_F$ - the average of all runs at the all factorial points
- $\bar{y}_C$ - the average of the n_C runs at the center point
- $\bar{y}_F - \bar{y}_C$ the difference between two averages

If the difference $\bar{y}_F - \bar{y}_C$ is small, then the center point lie on or near the hyperplane passing through the factorial points.

If the difference $\bar{y}_F - \bar{y}_C$ is large, then the quadratic curvature is present.

A single-degree-of-freedom **sum of squares for pure quadratic curvature**

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

SS can be incorporated into ANOVA and run hypothesis testing

$$H_0 : \sum_{j=1}^{k} \beta_{jj} = 0 \quad H_1 : \sum_{j=1}^{k} \beta_{jj} \neq 0.$$

Instead of ANOVA we can use a t-test to test for curvature (in estimation of regression coefficients in a linear model with).

Example from the last lesson - Unreplicated 2^k Factorial Designs

The 2^4 factorial design was used to investigate the effects of four factors on the filtration rate of a resin for a chemical process plant. The factors are:
A: temperature,
B: pressure,
C: concentration of chemical formaldehyde,
D: stirring rate.

Run Number	Factor A	B	C	D	Run Label	Filtration Rate (gal/h)
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

Unreplicated 2^k Factorial Designs

Pilot Plant Filtration Rate Experiment - FrF2 package

```
Factor settings:        Responses:
A  B  C  D           [1] Filtration
1 -1 -1 -1 -1
2  1  1  1  1

The design itself:
A  B  C  D Filtration    A  B  C  D Filtration
1  -1 -1 -1 -1         45   -1 -1 -1  1          43
2   1 -1 -1 -1         71    1 -1 -1  1         100
3  -1  1 -1 -1         48   -1  1 -1  1          45
4   1  1 -1 -1         65    1  1 -1  1         104
5  -1 -1  1 -1         68   -1 -1  1  1          75
6   1 -1  1 -1         60    1 -1  1  1          86
7  -1  1  1 -1         80   -1  1  1  1          70
8   1  1  1 -1         65    1  1  1  1          96
```

Pilot Plant Filtration Rate Experiment

Box plot

```
boxplot(Filtration ~ A , main = "Factor A - Temperature")
```

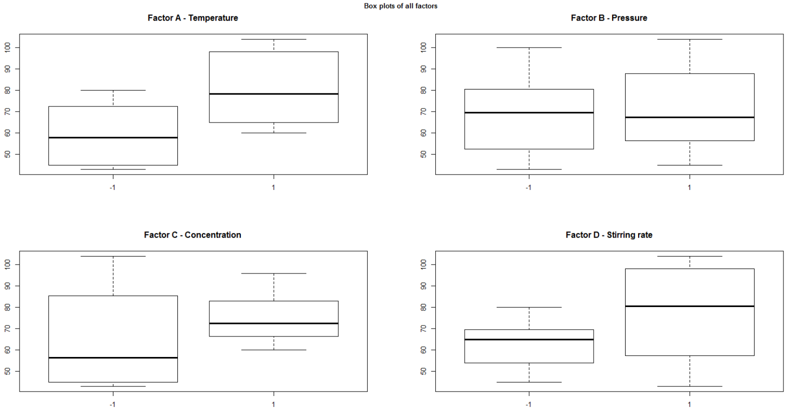


Illustration of center points adding in Pilot Plant Filtration Rate Experiment

Suppose that four center points are added to this experiment, and at the points $x_1 = x_2 = x_3 = x_4 = 0$ the four observed filtration rates were 73, 75, 66, 69.

The average of four center points is $\bar{y}_C = 70.75$, and the average of the 16 factorial runs is $\bar{y}_F = 70.06$. Since are very similar, we suspect that there is no strong curvature present.

$$MS_E = \frac{SS_E}{n_C - 1} = \frac{\sum_{Center\ points}(y_i - \bar{y})}{n_C - 1} = \frac{\sum_{i=1}^{4}(y_i - 70.75)^2}{4 - 1} = 16.25$$

$$SS_{Pure\ quadratic} = \frac{n_F n_C(\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} = \frac{(16)(4)(-0.69)^2}{16 + 4} = 1.51$$

Pure Error has 3 DF and Curvature has only 1 DF. The F-statistic is 0.093 and the P-value of the test is 0.780924.

We cannot reject the null hypothesis!

Addition of Center Points to a 2^k Designs

Anova tables for design with center points.

```
> rate3$E <- c(rep(0,times=16),rep(1,times=4))
> aov_model2= aov(Rate~A*B*C*D+E,data=rate3)
> summary(aov_model2)
Df Sum Sq Mean Sq F value  Pr(>F)
A             1 1870.6  1870.6 115.112 0.00173 **
B             1   39.1    39.1   2.404 0.21882
C             1  390.1   390.1  24.004 0.01627 *
D             1  855.6   855.6  52.650 0.00540 **
E             1    1.5     1.5   0.093 0.78024
A:B           1    0.1     0.1   0.004 0.95445
A:C           1 1314.1  1314.1  80.865 0.00290 **
B:C           1   22.6    22.6   1.388 0.32362
A:D           1 1105.6  1105.6  68.035 0.00373 **
B:D           1    0.6     0.6   0.035 0.86427
C:D           1    5.1     5.1   0.312 0.61569
A:B:C         1   14.1    14.1   0.865 0.42086
A:B:D         1   68.1    68.1   4.188 0.13320
A:C:D         1   10.6    10.6   0.650 0.47910
B:C:D         1   27.6    27.6   1.696 0.28376
A:B:C:D       1    7.6     7.6   0.465 0.54407
Residuals     3   48.8    16.2
```

$SS_{Purequadratic} = 1.51$ and $SS_{Pureerror} = 48.75$

Addition of Center Points to a 2^k Designs

Anova tables for original design and for design with center points.

```
> anova(aov(Filtration2~A*C+A*D,  data=rate2))
Response: Filtration2
          Df  Sum Sq Mean Sq F value    Pr(>F)
A          1 1870.56 1870.56  95.865 1.928e-06 ***
C          1  390.06  390.06  19.990  0.001195 **
D          1  855.56  855.56  43.847 5.915e-05 ***
A:C        1 1314.06 1314.06  67.345 9.414e-06 ***
A:D        1 1105.56 1105.56  56.659 1.999e-05 ***
Residuals 10  195.13   19.51

> anova(aov(Filtration4~A*C+A*D,  data=rate4))
Response: Filtration4
          Df  Sum Sq Mean Sq F value    Pr(>F)
A          1 1870.56 1870.56 106.721 6.235e-08 ***
C          1  390.06  390.06  22.254 0.0003301 ***
D          1  855.56  855.56  48.812 6.382e-06 ***
A:C        1 1314.06 1314.06  74.971 5.389e-07 ***
A:D        1 1105.56 1105.56  63.075 1.490e-06 ***
Residuals 14  245.39   17.53
```

$$SS_{Purequadratic} + SS_{LackofFit} + SS_{Pureerror} = 1.51 + 195.13 + 48.75 = 245.39$$

Addition of Center Points to a 2^k Designs

Is a quadratic effect needed (is the cube indificator significant)?

```
> summary(lm(Filtration4~A*C+A*D+iscube(rate4), rate4))
Call:
lm.default(formula = Filtration4 ~ A*C + A*D + iscube(rate4),
    data = rate4)

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        70.7500     2.1656  32.670 7.27e-14 ***
A                  10.8125     1.0828   9.986 1.83e-07 ***
C                   4.9375     1.0828   4.560 0.000535 ***
D                   7.3125     1.0828   6.753 1.36e-05 ***
iscube(rate4)TRUE  -0.6875     2.4212  -0.284 0.780924
A:C                -9.0625     1.0828  -8.369 1.36e-06 ***
A:D                 8.3125     1.0828   7.677 3.50e-06 ***

Residual standard error: 4.331 on 13 degrees of freedom
Multiple R-squared: 0.9578,     Adjusted R-squared: 0.9383
F-statistic:  49.2 on 6 and 13 DF,  p-value: 3.424e-08
```

The null hypothesis cannot be rejected.

Addition of Center Points to a 2^k Designs

Function `iscube` produce factor variable, generally we can add any quadratic variable by `I(Aˆ2)`:

```
lm.default(formula = Filtr ~ A + C + D + A:C + A:D + I(A^2),
data = rate_num2)
Residuals:
Min       1Q   Median       3Q      Max
-6.3750 -1.8750  0.0625  2.9062  5.7500
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  70.7500     2.1656  32.670 7.27e-14 ***
A            10.8125     1.0828   9.986 1.83e-07 ***
C             4.9375     1.0828   4.560 0.000535 ***
D             7.3125     1.0828   6.753 1.36e-05 ***
I(A^2)       -0.6875     2.4212  -0.284 0.780924
A:C          -9.0625     1.0828  -8.369 1.36e-06 ***
A:D           8.3125     1.0828   7.677 3.50e-06 ***
---
Signif. codes:  0 Â‘***Â’ 0.001 Â‘**Â’ 0.01 Â‘*Â’ 0.05 Â‘.Â’ 0
Residual standard error: 4.331 on 13 degrees of freedom
Multiple R-squared:  0.9578, Adjusted R-squared:  0.9383
F-statistic:  49.2 on 6 and 13 DF,  p-value: 3.424e-08
```

Blocking in the 2^k Factorial Design

Sometimes, it is not feasible or practical to run completely randomized experiment.

- Blocking is a technique for dealing with controllable nuisance variables (design factors that probably have an effect on the response, but we are not interested in that effect)
- Two cases are considered (Replicated designs x Unreplicated designs)
- If there are *n* replicates of the design, then each replicate is a block
- Each replicate is run in one of the blocks (time periods, batches of raw material, etc.)
- Runs within the block are randomized

Simple Blocking Example

Chemical Process Example

A - reactant concentration, B - catalyst amount, y - recovery
Nuisance Variable - Raw Materials. Only four trials per batch.

	Factor		Treatment	Replicate			
	A	*B*	Combination	I	II	III	Total
(1)	−	−	*A* low, *B* low	28	25	27	80
a	+	−	*A* high, *B* low	36	32	32	100
b	−	+	*A* low, *B* high	18	19	23	60
ab	+	+	*A* high, *B* high	31	30	29	90

Simple Blocking Example

2 factors, 3 replicates (blocks)

	Block 1	Block 2	Block 3
	$(1) = 28$	$(1) = 25$	$(1) = 27$
	$a = 36$	$a = 32$	$a = 32$
	$b = 18$	$b = 19$	$b = 23$
	$ab = 31$	$ab = 30$	$ab = 29$
Block totals:	$B_1 = 113$	$B_2 = 106$	$B_3 = 111$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Blocks	6.50	2	3.25		
A (concentration)	208.33	1	208.33	50.32	0.0004
B (catalyst)	75.00	1	75.00	18.12	0.0053
AB	8.33	1	8.33	2.01	0.2060
Error	24.84	6	4.14		
Total	323.00	11			

Confounding in the 2^k Factorial Design

What do we do if data for each combination of factor levels can not be collected under the same experimental conditions for an unreplicated 2^k design and the block size is smaller than the number of treatment combinations in one replicate?

- Consider the 2^5 case. May be impossible to replicate all in one block.
- The common blocking method for 2^k designs is to confound blocks with certain high order interactions.

How do we confound into two blocks?

- Choose highest order interaction.
- For this contrast, assign all $+$ to one block, and all $-$ to the other block.
- Randomize the order in which the experiments are run within the block.

Confounding in Pilot Plant Filtration Rate Problem

Number	A	B	C	D	Run Label	(gal/h)
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

Suppose only 8 runs can be made from one batch of raw material.

Confounding in Pilot Plant Filtration Rate Problem

Suppose only 8 runs can be made from one batch of raw material

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	$ABCD$
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
$abcd$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

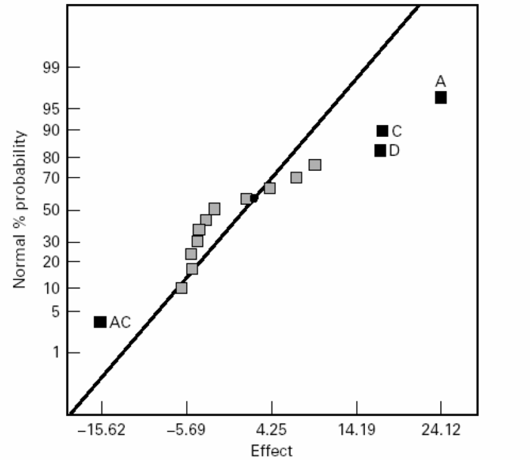
Example of confounding with ABCD interaction.
The ABCD interaction (or the block effect) will not considered as part of the error term.
The result of the analysis will be unchanged from mentioned example.

Confounding in Pilot Plant Filtration Rate Problem

Let the first eight runs (in run order) have filtration rate reduced by 20 units. This is the simulated "block effect".

Run Order	Std. Order	Factor A: Temperature	Factor B: Pressure	Factor C: Concentration	Factor D: Stirring Rate	Response Filtration Rate
8	1	−1	−1	−1	−1	25
11	2	1	−1	−1	−1	71
1	3	−1	1	−1	−1	28
3	4	1	1	−1	−1	45
9	5	−1	−1	1	−1	68
12	6	1	−1	1	−1	60
2	7	−1	1	1	−1	60
13	8	1	1	1	−1	65
7	9	−1	−1	−1	1	23
6	10	1	−1	−1	1	80
16	11	−1	1	−1	1	45
5	12	1	1	−1	1	84
14	13	−1	−1	1	1	75
15	14	1	−1	1	1	86
10	15	−1	1	1	1	70
4	16	1	1	1	1	76

Confounding in Pilot Plant Filtration Rate Problem



From the previous case one important interaction is not identified (AD).

Failing to block when we should have causes problems in interpretation the result of an experiment and can mask the presence of real factor effects.

Suggested Blocking Arrangements for the 2^k Factorial Design

Number of Factors, k	Number of Blocks, 2^p	Block Size, 2^{k-p}	Effects Chosen to Generate the Blocks	Interactions Confounded with Blocks
3	2	4	*ABC*	*ABC*
	4	2	*AB, AC*	*AB, AC, BC*
4	2	8	*ABCD*	*ABCD*
	4	4	*ABC, ACD*	*ABC, ACD, BD*
	8	2	*AB, BC, CD*	*AB, BC, CD, AC, BD, AD, ABCD*
5	2	16	*ABCDE*	*ABCDE*
	4	8	*ABC, CDE*	*ABC, CDE, ABDE*
	8	4	*ABE, BCE, CDE*	*ABE, BCE, CDE, AC, ABCD, BD, ADE*
	16	2	*AB, AC, CD, DE*	All two- and four-factor interactions (15 effects)
6	2	32	*ABCDEF*	*ABCDEF*
	4	16	*ABCF, CDEF*	*ABCF, CDEF, ABDE*
	8	8	*ABEF, ABCD, ACE*	*ABEF, ABCD, ACE, BCF, BDE, CDEF, ADF*
	16	4	*ABF, ACF, BDF, DEF*	*ABF, ACF, BDF, DEF, BC, ABCD, ABDE, AD, ACDE, CE, CDF, BCDEF, ABCEF, AEF, BE*
	32	2	*AB, BC, CD, DE, EF*	All two-, four-, and six-factor interactions (31 effects)
7	2	64	*ABCDEFG*	*ABCDEFG*
	4	32	*ABCFG, CDEFG*	*ABCFG, CDEFG, ABDE*
	8	16	*ABCD, CDEF, ADFG*	*ABC, DEF, AFG, ABCDEF, BCFG, ADEG, BCDEG*
	16	8	*ABCD, EFG, CDE, ADG*	*ABCD, EFG, CDE, ADG, ABCDEFG, ABE, BCG, CDFG, ADEF, ACEG, ABFG, BCEF, BDEG, ACF, BDF*

Blocking a Full 2^k into 2, 4, 8, $\ldots$ Blocks

Two blocks (size 2^{k-1} each): pick a *block generator* G (a high-order interaction column, e.g. $ABC\cdots$). Define block by $G = +1$ vs $G = -1$. Then:

- The **block effect is aliased with** G (that effect is not estimable).
- **All other effects orthogonal to G remain unconfounded with blocks**. Main effects and low-order interactions are typically safe if G is high order.
- Within each block, **each factor has both $+$ and $-$ levels** (balance holds because G is orthogonal to each main-effect column).

Four blocks (size 2^{k-2}): choose two independent generators G_1, G_2. Blocks are the 2^2 combinations of $(G_1, G_2) \in \{\pm 1\}^2$. Effects aliased with any G_i (or their products) are confounded with blocks.

2^b **blocks:** use b independent block generators; blocks are indexed by their signs.

Blocking with more than two block

- The two level factorial can be confounded in $2, 4, 8, \ldots (2p,\ p > 1)$ blocks.
- For four blocks, select two effects to confound, automatically confounding a third effect.
- Choice of confounding schemes non trivial - see enclosed paper.

Blocking - summary, again

- Block when you can and randomize what you cannot.
- When in doubt, block.
- Block out the nuisance variables you know about, randomize as much as possible and rely on randomization to help balance out unknown nuisance effects.

Exercises

Exercise:
Finish problems 6.31 and 6.32 from the last lecture.

Exercise:
Solve problems 6.26, 6.27 following by 7.7 and 7.8. described below (from C. Montgomery DAoE - 8. edition).

An experiment was run in a semiconductor fabrication plant in an effort to increase yield. Five factors, each at two levels, were studied. The factors (and levels) were A = aperture setting (small, large), B = exposure time (20% below nominal, 20% above nominal), C = development time (30 and 45 s), D = mask dimension (small, large), and E = etch time (14.5 and 15.5 min). The unreplicated 2^5 design was run.

$(1) = 7$	$d = 8$	$e = 8$	$de = 6$
$a = 9$	$ad = 10$	$ae = 12$	$ade = 10$
$b = 34$	$bd = 32$	$be = 35$	$bde = 30$
$ab = 55$	$abd = 50$	$abe = 52$	$abde = 53$
$c = 16$	$cd = 18$	$ce = 15$	$cde = 15$
$ac = 20$	$acd = 21$	$ace = 22$	$acde = 20$
$bc = 40$	$bcd = 44$	$bce = 45$	$bcde = 41$
$abc = 60$	$abcd = 61$	$abce = 65$	$abcde = 63$

Exercises

1. Construct a normal probability plot of the effect estimates. Which effects appear to be large?
2. Conduct an analysis of variance to confirm your findings for part (1).
3. Write down the regression model relating yield to the significant process variables.
4. Plot the residuals on normal probability paper. Is the plot satisfactory?
5. Plot the residuals versus the predicted yields and versus each of the five factors. Comment on the plots.
6. Interpret any significant interactions.
7. What are your recommendations regarding process operating conditions?
8. Project the 2^5 design in this problem into a 2^k design in the important factors. Sketch the design and show the average and range of yields at each run. Does this sketch aid in interpreting the results of this experiment?

Exercises

9. Suppose that the experimenter had run four center points in addition to the 32 trials in the original experiment. The yields obtained at the center point runs were 68, 74, 76, and 70. Reanalyze the experiment, including a test for pure quadratic curvature. Discuss what your next step would be.
10. Construct and analyze a design in two blocks with ABCDE confounded with blocks.
11. Construct and analyze a design in four blocks. Suggest a reasonable confounding scheme.