Purdue-NCKU program

Lecture 6 2^k Factorial Design

Dr. Qifan Song

2^k Factorial Design

- Involving k factors
- Each factor has two levels (often labeled + and -)
- Factor screening experiment (preliminary study)
- Factors need not be on numeric scale
- Identify important factors and their interactions

• Interaction (of any order) has **ONE** degree of freedom

2² Factorial Design

Example:

fac	tor		re	plica	te	
А	В	treatment	1	2	3	mean
		(1)	28	25	27	80/3
+	—	а	36	32	32	100/3
	+	b	18	19	23	60/3
+	+	ab	31	30	29	90/3

- $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$
- Let \$\overline{y}(A_+)\$, \$\overline{y}(A_-)\$, \$\overline{y}(B_+)\$ and \$\overline{y}(B_-)\$ be the level means of A and B.
- Let $\bar{y}(A_-B_-)$, $\bar{y}(A_+B_-)$, $\bar{y}(A_-B_+)$ and $\bar{y}(A_+B_+)$ be the treatment means

Main Effect

Define main effects of A (denoted again by A) as follows:

$$A = m.e.(A) = \bar{y}(A_{+}) - \bar{y}(A_{-})$$

= $\frac{1}{2}(\bar{y}(A_{+}B_{+}) + \bar{y}(A_{+}B_{-})) - \frac{1}{2}(\bar{y}(A_{-}B_{+}) + \bar{y}(A_{-}B_{-})))$
= $\frac{1}{2}(\bar{y}(A_{+}B_{+}) + \bar{y}(A_{+}B_{-}) - \bar{y}(A_{-}B_{+}) - \bar{y}(A_{-}B_{-})))$
= $\frac{1}{2}(-\bar{y}(A_{-}B_{-}) + \bar{y}(A_{+}B_{-}) - \bar{y}(A_{-}B_{+}) + \bar{y}(A_{+}B_{+})))$
= 8.33

• Let $C_A = (-1, 1, -1, 1)$, a contrast on treatment mean responses, then

m.e.(A)
$$= \frac{1}{2} \hat{C}_A$$

Notice that

$$A = m.e.(A) = (\bar{y}(A_{+}) - \bar{y}_{..}) - (\bar{y}(A_{-}) - \bar{y}_{..}) = \hat{\tau}_{2} - \hat{\tau}_{1} = 2\hat{\tau}_{2}$$

Main effect is defined in a different way than the factorial modeling. But they are connected and equivalent.

Interaction

- Interaction between A and B: does the effect of A depend on the level of B?
- Define interaction between A and B

$$AB = \text{Int}(AB) = \frac{1}{2}(m.e.(A \mid B_{+}) - m.e.(A \mid B_{-}))$$

$$= \frac{1}{2}(\bar{y}(A_{+} | B_{+}) - \bar{y}(A_{-} | B_{+})) - \frac{1}{2}(\bar{y}(A_{+} | B_{-}) - \bar{y}(A_{-} | B_{-})))$$

$$= \frac{1}{2}(\bar{y}(A_{-}B_{-}) - \bar{y}(A_{+}B_{-}) - \bar{y}(A_{-}B_{+}) + \bar{y}(A_{+}B_{+}))$$

Let $C_{AB} = (1, -1, -1, 1)$, a contrast on treatment means, then

$$AB = Int(AB) = \frac{1}{2}\hat{C}_{AB}$$

• Notice that $Int(AB) = \hat{\tau \beta}_{22} - \hat{\tau \beta}_{21} = similar$ interaction factorial effects difference $=2\hat{\tau \beta}_{22}$

Effects and Contrasts

fac	tor			ef	fect	(cor	itrast)
А	В	total	mean	Ι	А	В	AB
_		80	80/3	1	-1	-1	1
+		100	100/3	1	1	-1	-1
_	+	60	60/3	1	-1	1	-1
+	+	90	90/3	1	1	1	1

- There is a one-to-one correspondence between effects and contrasts, and contrasts can be directly used to estimate the effects.
- For a effect corresponding to contrast $c = (c_1, c_2, ...)$ in 2^2 design

$$\text{effect} = \frac{1}{2} \sum_{i} c_i \bar{y}_i$$

where i is an index for treatments and the summation is over all treatments.

• Pay attention to the column of the contrast matrix

2³ Factorial Design

f	actc	or		resp	onse	
А	В	С	treatment	1	2	total
_	—		(1)	-3	-1	-4
+			а	0	1	1
—	+		b	-1	0	-1
+	+		ab	2	3	5
—	—	+	С	-1	0	-1
+		+	ас	2	1	3
—	+	+	bc	1	1	2
+	+	+	abc	6	5	11

 $y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\tau\gamma)_{ik} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$

factorial effects and contrasts

Main effects:

$$A = m.e.(A) = \bar{y}(A_{+}) - \bar{y}(A_{-})$$

= $\frac{1}{4}(\bar{y}(---) + \bar{y}(+--) - \bar{y}(-+-) + \bar{y}(++-) - \bar{y}(--+))$
+ $\bar{y}(+-+) - \bar{y}(-++) + \bar{y}(+++))$
= $3.00 = 2\hat{\tau}_2$
The contrast is (-1,1,-1,1,-1,1)

$$B: (-1, -1, 1, 1, -1, -1, 1, 1), B = 2.25$$

$$C: (-1, -1, -1, -1, 1, 1, 1), C = 1.75$$

2-factor interactions:

AB: $A \times B$ componentwise, $AB = .75 = \hat{\tau}\beta_{22}$

AC: $A \times C$ componentwise, AC=.25

BC: $B \times C$ componentwise, BC=.50

High order interaction

k-th order interaction means: does the (k - 1)-th interaction depend on level of the k-th factor

3-factor interaction:

$$ABC = int(ABC) = \frac{1}{2}(int(AB \mid C+) - int(AB \mid C-))$$

= $\frac{1}{4}(-\bar{y}(--) + \bar{y}(+-) + \bar{y}(-+) - \bar{y}(++-))$
+ $\bar{y}(--+) - \bar{y}(+-+) - \bar{y}(-++) + \bar{y}(+++))$
= $.50 = 2\tau \hat{\beta}\gamma_{222}$

The contrast is $(-1, 1, 1, -1, -1, -1, 1) = A \times B \times C$.

Contrasts for Calculating Effects in 2³ Design

				factorial effects							
А	В	С	treatment	Ι	A	B	AB	C	AC	BC	ABC
		—	(1)	1	-1	-1	1	-1	1	1	-1
+		—	а	1	1	-1	-1	-1	-1	1	1
_	+	—	b	1	-1	1	-1	-1	1	-1	1
+	+	_	ab	1	1	1	1	-1	-1	-1	-1
_		+	С	1	-1	-1	1	1	-1	-1	1
+		+	ас	1	1	-1	-1	1	1	-1	-1
_	+	+	bc	1	-1	1	-1	1	-1	1	-1
+	+	+	abc	1	1	1	1	1	1	1	1

Estimates:

grand mean:
$$\frac{\sum \bar{y}_{i.}}{2^3}$$

effect : $\frac{\sum c_i \bar{y}_{i.}}{2^{3-1}}$

General 2^k Design

- k factors: A, B, ..., K each with 2 levels (+,-)
- consists of all possible level combinations (2 k treatments) each with n replicates
- Classify factorial effects:

type of effect	label	the number of effects
main effects (of order 1)	A, B, C,, K	k
2-factor interactions (of order 2)	AB, AC ,, JK	$\left(\begin{array}{c}k\\2\end{array}\right)$
3-factor interactions (of order 3)	ABC, ABD, \ldots, IJK	$\left(\begin{array}{c}k\\3\end{array}\right)$
•••	•••	••••
k-factor interaction (of order k)	$ABC \cdots K$	$\left(egin{array}{c} k \\ k \end{array} ight)$

- Each effect (main or interaction) has 1 degree of freedom full model (i.e. model consisting of all the effects) has 2^k-1 degrees of freedom.
- Error component has $2^k(n-1)$ degrees of freedom.
- One-to-one correspondence between effects and contrasts:
 - For main effect: convert the level column of a factor using $\Rightarrow -1$ and $+ \Rightarrow 1$
 - For interactions: multiply the contrasts of the main effects of the involved factors, componentwisely.

General 2^k Design: Analysis

• Estimates:

grand mean : $\frac{\sum \bar{y}_{i.}}{2^k}$ For effect with contrast $C = (c_1, c_2, \dots, c_{2^k})$, its estimate is

effect =
$$\frac{\sum c_i \bar{y}_i}{2^{(k-1)}}$$

• Variance

Var(effect) =
$$\frac{\sigma^2}{n2^{k-2}}$$

S.E.(effect) = $\frac{MSE}{n2^{k-2}}$

• C.I. for every factorial effect

$$\mathsf{effect} \pm t_{\alpha/2, 2^k(n-1)}\mathsf{S.E.}(\mathsf{effect})$$

Unreplicated 2^k Design

- *n* = 1
- No degree of freedom left for error component if full model is fitted.
- Same estimation method
- No error sum of squares available, cannot estimate σ^2 and test effects in both the ANOVA and Regression approaches.
- Approach 1: pooling high-order interactions
 - Often assume 3 or higher interactions do not occur
 - Pool estimates together for error
 - Warning: may pool significant interaction

• Approach 2: Using the normal probability plot (QQ plot) to identify significant effects.

– Recall

$$Var(effect) = \frac{\sigma^2}{2^{(k-2)}}$$

If the effect is not significant (=0), then the effect estimate follows $N(0, \frac{\sigma^2}{2^{(k-2)}})$

- Assume all effects not significant, their estimates can be considered as a random sample from $N(0, \frac{\sigma^2}{2(k-2)})$
- QQ plot of the estimates is expected to be a linear line
- Deviation from a linear line indicates significant effects

A case study

	tac	tor		
A	B	C	D	filtration
		_	_	45
+		_	—	71
_	+		_	48
+	+		_	65
_	_	+	_	68
+	_	+	_	60
_	+	+	_	80
+	+	+	_	65
			+	43
+	_	_	+	100
_	+		+	45
+	+		+	104
		+	+	75
+		+	+	86
_	+	+	+	70
+	+	+	+	96

ALL Ranked Effects

Obs	_NAME_	COL1	effect	neff
1	AC	-9.0625	-18.125	-1.73938
2	BCD	-1.3125	-2.625	-1.24505
3	ACD	-0.8125	-1.625	-0.94578
4	CD	-0.5625	-1.125	-0.71370
5	BD	-0.1875	-0.375	-0.51499
6	AB	0.0625	0.125	-0.33489
7	ABCD	0.6875	1.375	-0.16512
8	ABC	0.9375	1.875	-0.00000
9	BC	1.1875	2.375	0.16512
10	В	1.5625	3.125	0.33489
11	ABD	2.0625	4.125	0.51499
12	С	4.9375	9.875	0.71370
13	D	7.3125	14.625	0.94578
14	AD	8.3125	16.625	1.24505
15	Α	10.8125	21.625	1.73938



Effect Selection and Analysis

- Potentially significant effects: A, AD, C, D, AC.
- ANOVA model involving only A, C, D and their interactions (projecting the original unreplicated 2⁴ experiment onto a replicated 2³ experiment)
- Make inferences with non-ZERO MSE
- Diagnostics using residuals.

2^{k-p} Fractional Factorial Design

Fundamental Principles Regarding Factorial Effects

Suppose there are k factors (A, B, ..., J, K) in an experiment. All possible factorial effects include

effects of order 1: A, B, ..., K (main effects) effects of order 2: AB, AC,,JK (2-factor interactions)

• 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 (Pareto principle)
 - The number of relatively important effects in a factorial experiment is small
- Effect Heredity Principle
 - In order for an interaction to be significant, at least one of its parent factors should be significant.

Fractional Factorials

- May not have sources (time, money, etc) for full factorial design
- Number of runs required for full factorial grows quickly
 - Consider 2^k design
 - If $k = 7 \rightarrow 128$ runs required
 - Can estimate 127 effects
 - Only 7 df for main effects, 21 for 2-factor interactions
 - the remaining 99 df are for interactions of order ≥ 3
- Often only lower order effects are important
- Full factorial design may not be necessary according to
 - Hierarchical ordering principle
 - Effect Sparsity Principle
- A fraction of the full factorial design (i.e. a subset of all possible level combinations) is sufficient.

There are four factors in the experiment(A, B, C and D), each of 2 levels. Suppose the available resource is enough for conducting 8 runs. 2^4 full factorial design consists of all the 16 level combinations of the four factors. We need to choose half of them.

- If you drop one factors for a 2³ full factorial design, this factor and their interactions with other factors cannot be investigated.
- Want investigate all 4 factors in the experiment
- A fraction of 2⁴ full factorial design will be used.
- Confounding (aliasing) will happen because using a subset

The chosen half is called 2^{4-1} fractional factorial design.

2⁴⁻¹ Fractional Factorial Design

- the number of factors: k = 4
- the fraction index: p = 1
- the number of runs (level combinations): $N = \frac{2^4}{2^1} = 8$
- Construct 2⁴⁻¹ designs via "confounding" (aliasing)
 - select 3 factors (e.g. A, B, C) to form a 2^3 full factorial (basic design)
 - confound (alias) D with a high order interaction of A, B and C. For example,

		fact	orial	effect	ts (co	ntrast	ts)
Ι	Α	В	С	AB	AC	BC	ABC=D
1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1	-1
1	1	1	1	1	1	1	1

• Note: 1 corresponds to + and -1 corresponds to -.

Verify:

- 1. the chosen level combinations form a half of the 2^4 design.
- 2. the product of columns A, B, C and D equals 1, i.e.,

I = ABCD

which is called the **defining relation**, or *ABCD* is called a **defining word** (contrast).

Aliasing in 2^{4-1} Design

For four factors A, B, C and D, there are $2^4 - 1$ effects: A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD

Response	Ι	А	В	С	D	AB	 CD	ABC	BCD	 ABCD
y_1	1	-1	-1	-1	-1	1	 1	-1	-1	 1
y_2	1	1	-1	-1	1	-1	 -1	1	1	 1
y_3	1	-1	1	-1	1	-1	 -1	1	-1	 1
y_4	1	1	1	-1	-1	1	 1	-1	1	 1
y_5	1	-1	-1	1	1	1	 1	1	-1	 1
y_6	1	1	-1	1	-1	-1	 -1	-1	1	 1
y_7	1	-1	1	1	-1	-1	 -1	-1	-1	 1
y_8	1	1	1	1	1	1	 1	1	1	 1

Contrasts for main effects by converting - to -1 and + to 1; contrasts for other effects obtained by multiplication.

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)$$

$$BCD = \frac{1}{4}(-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8)$$

A, BCD are aliases or aliased. The contrast is for A+BCD. A and BCD are not distinguishable.

 $AB = \bar{y}_{AB+} - \bar{y}_{AB-} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8) \ CD = \bar{y}_{CD+} - \bar{y}_{CD-} = \frac{1}{4}(y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8)$

AB, CD are aliases or aliased. The contrast is for AB+CD. AB and CD are not distinguishable.

There are other 5 pairs. They are caused by the defining relation

I = ABCD,

that is, I (the intercept) and 4-factor interaction ABCD are aliased.

<u>Alias Structure for 2^{4-1} with I = ABCD</u>

- Alias Structure: I = ABCD A = A * I = A * ABCD = BCD $B = \dots = ACD$ $C = \dots = ABD$ $D = \dots = ABC$ AB = AB * I = AB * ABCD = CD $AC = \dots = BD$ $AD = \dots = BC$
- all 16 factorial effects for A, B, C and D are partitioned into 8 groups each with 2 aliased effects.
- When a low order effect is aliased with a high order effect, by Hierarchical Order principle, we tend to believe that the effect is mostly contributed by the low order effect

A Different 2^{4-1} Design

the defining relation I = ABD generates another 2⁴⁻¹ fractional factorial design, denoted by d₂. Its alias structure is given below.
 I = ABD
 A = BD

A = BDB = ADC = ABCDD = ABABC = CDACD = BCBCD = AC

• Recall d_1 is defined by I = ABCD. Comparing d_1 and d_2 , which one we should choose or which one is better?

1. Length of a defining word is defined to be the number of the involved factors.

2. **Resolution** of a fractioanl factorial design is defined to be the minimum length of the defining words, usually denoted by Roman numbers, III, IV, V, etc...

Resolution and Maximum Resolution Criterion

- d_1 : I = ABCD is a resolution IV design denoted by 2_{IV}^{4-1} .
- d_2 : I = ABD is a resolution III design denoted by 2_{III}^{4-1} .
- If a design is of resolution R, then none of the *i*-factor interactions is aliased with any other interaction of order less than R i.

 d_1 : main effects are not aliased with other main effects and 2-factor interactions

 d_2 : main effects are not aliased with main effects

- d_1 is better, because d_1 has higher resolution than d_2 . In fact, d_1 is optimal among all the possible fractional factorial 2^{4-1} designs
- Maximum Resolution Criterion fractional factorial design with maximum resolution is optimal

How to Analyze 2^{4-1} design

- Compute all effects
- Use QQ plot to determine which ones are significant
- Resolve the ambiguities in aliased effects via the fundamental principles beneficial
- Project the design to a replicated factorial design
- Example

- I=ABCD

- QQ plot determine A, B, CD are significant
- By HO principle C, D are not significant
- By EH principle, CD are not significant
- All significant effects are A, B and AB
- View the data as a 2^2 experiment with 2 replications

General 2^{k-1} Design

- k factors: A, B, ..., K
- can only afford half of all the combinations (2^{k-1})
- Basic design: a 2^{k-1} full factorial for k-1 factors: A, B, ..., J.
- The setting of kth factor is determined by alasing K with the ABC....J, i.e., $K = ABC \cdots J$
- Defining relation: $I = ABCD....\tilde{I}JK$. Resolution=k
- 2^k factorial effects are partitioned into 2^{k-1} groups each with two aliased effects.
- Use fundamental principles, domain knowledge, follow-up experiment to de-alias.

One Quarter Fraction: 2^{k-2} **Design**

Parts manufactured in an injection molding process are showing excessive shrinkage. A quality improvement team has decided to use a designed experiment to study the injection molding process so that shrinkage can be reduced. The team decides to investigate six factors

- A: mold temperature
- *B*: screw speed
- C: holding time
- D: cycle time
- E: gate size
- *F*: holding pressure

each at two levels, with the objective of learning about main effects and interactions.

They decide to use 16-run fractional factorial design.

- a full factorial has $2^6 = 64$ runs.
- 16-run is one quarter of the full factorial
- How to construct the fraction?

Injection Molding Experiment: 2^{6-2} Design

ba	asic	desig	gn			
A	B	C	D	E = ABC	F = BCD	shrinkage
	_			_	_	6
+	_	_	_	+	—	10
_	+	_	_	+	+	32
+	+	—		_	+	60
_	—	+	_	+	+	4
+	—	+		_	+	15
_	+	+		_	_	26
+	+	+		+	_	60
_	—		+	_	+	8
+	—	_	+	+	+	12
	+	—	+	+	<u> </u>	34
+	+	—	+	_	_	60
	_	+	+	+	_	16
+	_	+	+		_	5
	+	+	+	_	+	37
+	+	+	+	+	+	52

Two defining relations are used to generate the columns for E and F.

$$I = ABCE$$
, and $I = BCDF$

They induce another defining relation:

$$I = ABCE * BCDF = AB^2C^2DEF = ADEF$$

The complete defining relation:

$$I = ABCE = BCDF = ADEF$$

Defining contrasts subgroup: $\{I, ABCE, BCDF, ADEF\}$

Alias Structure

I = ABCE = BCDF = ADEF implies

$$A = BCE = ABCDF = DEF$$

Similarly, we can derive the other groups of aliased effects.

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

Wordlength pattern $W = (W_0, W_1, \dots, W_6)$, where W_i is the number of defining words of length *i* (i.e., involving *i* factors)

$$W = (1, 0, 0, 0, 3, 0, 0)$$

Resolution is the smallest *i* such that i > 0 and $W_i > 0$. Hence it is a 2_{IV}^{6-2} design

2^{6–2} Design: an Alternative

- Basic Design: A, B, C, D
- E = ABCD, F = ABC, i.e., I = ABCDE, and I = ABCF
- which induces: I = DEF
- complete defining relation: I = ABCDE = ABCF = DEF
- wordlength pattern: W = (1, 0, 0, 1, 1, 1, 0)

• Alias structure (ignore effects of order 3 or higher)

$A = \dots$	$AB = CF = \dots$
$B = \dots$	$AC = BF = \dots$
$C = \dots$	$AD = \dots$
$D = EF = \dots$	$AE = \dots$
$E = DF = \dots$	$AF = BC = \dots$
$F = DE = \dots$	BD =
	BE =
	CD =
	CE =

- an effect is said to be **clearly estimable** if it is not aliased with main effect or two-factor interactions.
- Which design is better d₁ or d₂? d₁ has six clearly estimable main effects while d₂ has three clearly estimable main effects and six clearly estimable two-factor ints.

Minimum Aberration Criterion

Recall 2^{k-p} with maximum resolution should be preferred. But, it is possible that there are two designs that attain the maximum resolution. How should we further distinguish them?

For example, consider 2^{7-2} fractional factorial design

*d*₁: basic design: *A*, *B*, *C*, *D*, *E*; F = ABC, G = ABDEcomplete defining relation: I = ABCF = ABDEG = CDEFGwordlength pattern: W = (1, 0, 0, 0, 1, 2, 0, 0)Resolution: IV

 d_2 : basic design: A, B, C, D, E; F = ABC, G = ADEcomplete defining relation: I = ABCF = ADEG = BCDEFGwordlength pattern: W = (1, 0, 0, 0, 2, 0, 1, 0)Resolution: IV

 d_1 and d_2 , which is better?

Minimum Aberration Criterion

Definition: Let d_1 and d_2 be two 2^{k-p} designs, let r be the smallest **positive** integer such that $W_r(d_1) \neq W_r(d_2)$.

If $W_r(d_1) < W_r(d_2)$, then d_1 is said to have less aberration than d_2 .

If there does not exist any other design that has less aberration than d_1 , then d_1 has minimum aberration.

Chapter Review

- 2^k design
- 2^k design without replication
- 2^{k-p} design