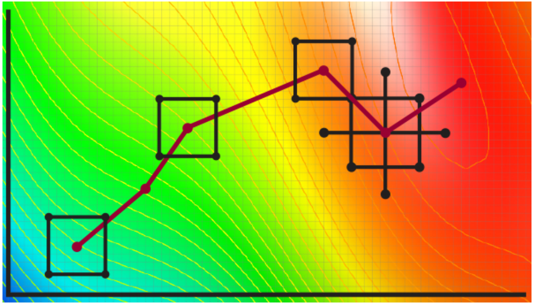
Experimentation for Improvement



http://learnche.org/

Design and Analysis of Experiments

Copyright, sharing, and attribution notice

Cell-culture example: creating the fractional factorial design

Experiment	A	B	C	D = AB	E = AC
1	−	−	−	+	+
2	+	−	−	−	−
3	−	+	−	−	+
4	+	+	−	+	−
5	−	−	+	+	−
6	+	−	+	−	+
7	−	+	+	−	−
8	+	+	+	+	+


[Flickr: londonmatt]

1. Read the generators from the trade off table
   - **D = AB** and **E = AC**
2. Rearrange the generators as **I = ...**
   - **I = ABD** and **I = ACE**
3. Form the **defining relationship** taking all combinations of the words: **I = ...**
   - **I = ABD = ACE = BCDE**
4. Ensure the defining relationship has 2^p words
   - $p = 2$, and we have 4 words.
5. **In this example:** we will use the defining relationship to compute the aliasing

The purpose of the defining relationship: to calculate all possible aliases

By example: what are the aliases (confounding) with factor **B**?

$$
\begin{aligned}
\mathbf{I\,B} &= \mathbf{ABD\,B} = \mathbf{ACE\,B} = \mathbf{BCDE\,B} \\
\mathbf{B} &= \mathbf{A(BB)D} = \mathbf{ABCE} = \mathbf{(BB)CDE} \\
\mathbf{B} &= \mathbf{AID} = \mathbf{ABCE} = \mathbf{ICDE} \\
\mathbf{B} &= \mathbf{AD} = \mathbf{ABCE} = \mathbf{CDE}
\end{aligned}
$$

- Take the defining relationship
- Multiply every word by a **B** term
- Rearrange the order
- Simplify by using the rules: **AA=I**; **BB=I**; ... **DD=I**, *etc*
- Drop out the unnecessary identity terms

How do we use this aliasing information we obtain?

Now we know the aliases of **B**:

$$\mathbf{B} \quad = \quad \mathbf{AD} \quad = \quad \mathbf{ABCE} \quad = \quad \mathbf{CDE}$$

- These are the terms that B is going to be confounded with
- We cannot tell B apart from the AD interaction.

How do we use this aliasing information we obtain?

Now we know the aliases of **B**:

$$\mathbf{B} \quad = \quad \mathbf{AD} \quad = \quad \mathbf{ABCE} \quad = \quad \mathbf{CDE}$$

- These are the terms that B is going to be confounded with
- We cannot tell B apart from the AD interaction.

How do we use this aliasing information we obtain?

Try this yourself: what are the aliases of **C**?

$$\mathbf{C} = \mathbf{ABCD} = \mathbf{AE} = \mathbf{BDE}$$

- These are the terms that C is going to be confounded with
- We cannot tell C apart from the AE interaction.

How do we use this aliasing information we obtain?

Try this yourself: what are the aliases of **A**?

$$\mathbf{A} = \mathbf{BD} = \mathbf{CE} = \mathbf{ABCDE}$$

- These are the terms that A is going to be confounded with
- We cannot tell A apart from the BD interaction and the CE interaction.

How do we use this aliasing information we obtain?

Write out all the aliases for the 5 main effects:

$$
\begin{array}{ccccccc}
\mathbf{A} & = & \mathbf{BD} & = & \mathbf{CE} & = & \mathbf{ABCDE} \\
\mathbf{B} & = & \mathbf{AD} & = & \mathbf{ABCE} & = & \mathbf{CDE} \\
\mathbf{C} & = & \mathbf{ABCD} & = & \mathbf{AE} & = & \mathbf{BDE} \\
\mathbf{D} & = & \mathbf{AB} & = & \mathbf{ACDE} & = & \mathbf{BCE} \\
\mathbf{E} & = & \mathbf{ABDE} & = & \mathbf{AC} & = & \mathbf{BCD}
\end{array}
$$

How do we use this aliasing information we obtain?

Aliases for the 5 main effects (dropping out 3rd order and higher interactions):

$$
\begin{aligned}
\mathbf{A} &= \mathbf{BD} = \mathbf{CE} \\
\mathbf{B} &= \mathbf{AD} \\
\mathbf{C} &= \mathbf{AE} \\
\mathbf{D} &= \mathbf{AB} \\
\mathbf{E} &= \mathbf{AC}
\end{aligned}
$$

Back to some more baking experiments

Recipe: http://yint.org/honeycomb-cake

[Flickr: andrea_nguyen]

Potential factors to consider are

- stirring speed
- type of coconut milk
- baking soda
- amount of wheat starch
- baking temperature

Aliasing pattern for a 2^{5-2}_{III} design:

$$
\begin{aligned}
\mathbf{A} &= \mathbf{BD} = \mathbf{CE}\\
\mathbf{B} &= \mathbf{AD}\\
\mathbf{C} &= \mathbf{AE}\\
\mathbf{D} &= \mathbf{AB}\\
\mathbf{E} &= \mathbf{AC}
\end{aligned}
$$

Back to some more baking experiments

Recipe: http://yint.org/honeycomb-cake

[Flickr: andrea_nguyen]

So my factor assignment so far is:

- stirring speed
- type of coconut milk
- baking soda
- amount of wheat starch
- **A** = baking temperature

Aliasing pattern for a 2^{5-2}_{III} design:

$$\begin{array}{ccccc} \mathbf{A} & = & \mathbf{BD} & = & \mathbf{CE} \\ \mathbf{B} & = & \mathbf{AD} & & \\ \mathbf{C} & = & \mathbf{AE} & & \\ \mathbf{D} & = & \mathbf{AB} & & \\ \mathbf{E} & = & \mathbf{AC} & & \end{array}$$

Back to some more baking experiments

Recipe: http://yint.org/honeycomb-cake

[Flickr: andrea_nguyen]

So my factor assignment so far is:

- **B** = stirring speed (to avoid an AB interaction)
- type of coconut milk
- baking soda
- amount of wheat starch
- **A** = baking temperature

Aliasing pattern for a 2^{5-2}_{III} design:

$$\begin{array}{ccccc} \mathbf{A} & = & \mathbf{BD} & = & \mathbf{CE} \\ \mathbf{B} & = & \mathbf{AD} & & \\ \mathbf{C} & = & \mathbf{AE} & & \\ \mathbf{D} & = & \mathbf{AB} & & \\ \mathbf{E} & = & \mathbf{AC} & & \end{array}$$

Back to some more baking experiments

Recipe: http://yint.org/honeycomb-cake

[Flickr: andrea_nguyen]

So my factor assignment so far is:

- **B** = stirring speed (to avoid an AB interaction)
- type of coconut milk
- **D** = baking soda (we expect this to be sensitive)
- amount of wheat starch
- **A** = baking temperature

Aliasing pattern for a 2^{5-2}_{III} design:

$$\begin{array}{lllll}
\mathbf{A} & = & \mathbf{BD} & = & \mathbf{CE} \\
\mathbf{B} & = & \mathbf{AD} & & \\
\mathbf{C} & = & \mathbf{AE} & & \\
\mathbf{D} & = & \cancelto{0}{\mathbf{AB}} & & \text{(to get an unbiased estimate)} \\
\mathbf{E} & = & \mathbf{AC} & &
\end{array}$$

How do we use this aliasing information we obtain?

The aliases for the 5 main effects in this a 2^{5-2}_{III} design:

$$
\begin{array}{ccccccc}
\mathbf{A} & = & \mathbf{BD} & = & \mathbf{CE} & = & \mathbf{ABCDE} \\
\mathbf{B} & = & \mathbf{AD} & = & \mathbf{CDE} & = & \mathbf{ABCE} \\
\mathbf{C} & = & \mathbf{AE} & = & \mathbf{BDE} & = & \mathbf{ABCD} \\
\mathbf{D} & = & \mathbf{AB} & = & \mathbf{BCE} & = & \mathbf{ACDE} \\
\mathbf{E} & = & \mathbf{AC} & = & \mathbf{BCD} & = & \mathbf{ABDE}
\end{array}
$$

Is is possible to get a more favourable fractional factorial?

In other words, can we get better confounding?

- In the 5-factor example:

  main effects are confounded with two factor interactions

Is it possible to achieve a design where main effects are confounded with 3-factor interactions?

These are called resolution IV designs.

Example to try yourself: find the defining relationship

You'd like to investigate 6 factors, and have a budget for 15 to 20 experiments.

1. Read the generators from the table
2. Rearrange the generators as $\mathbf{I} = \ldots$
3. Form the defining relationship taking all combinations of the words, so that $\mathbf{I} = \ldots$
4. Ensure the defining relationship has 2^p words
5. Use this defining relationship to compute the aliasing pattern

Example for practice: solution for finding the defining relationship

You'd like to investigate 6 factors, and have a budget for 15 to 20 experiments.

1. Read the generators from the table for $k = 6$, and 16 experiments
   - **E = ABC** and **F = BCD**
2. Rearrange the generators as **I = . . .**
   - **I = ABCE** and **I = BCDF**
3. Form the defining relationship taking all combinations of the words, so that **I = . . .**
   - **I = ABCE = BCDF = A(BB)(CC)DEF = ADEF**
4. Ensure the defining relationship has 2^p words
   - It does; since $p = 2$ in this case
5. We will use this defining relationship now

$2^{6\text{-}2}_{\text{IV}}$

±E=ABC
±F=BCD

The 6-factor example in 16 experiments: a resolution IV design

By example: what are the aliases (confounding) with factor **A**?

The defining relationship is

$$\mathbf{I} = \mathbf{ABCE} = \mathbf{BCDF} = \mathbf{ADEF}$$

Now multiply all words by **A**:

$$\mathbf{A} = \mathbf{BCE} = \mathbf{ABCDF} = \mathbf{DEF}$$

There are only 3-factor and higher level interactions*!*

The 6-factor example in 16 experiments: a resolution IV design

What are the aliases (confounding) with a two-factor interaction term: **CD**?

The defining relationship is

$$\mathbf{I} = \mathbf{ABCE} = \mathbf{BCDF} = \mathbf{ADEF}$$

Now multiply all words by **CD**:

$$\mathbf{CD} = \mathbf{ABDE} = \mathbf{BF} = \mathbf{ACEF}$$

Two factor interactions are aliased with other two factor interactions.

increasing cost

Number of factors, *k*

Number of runs	3	4	5	6	7	8	9
4	2^{3-1}_{III} ±C=AB						
8	2^{3} *full*	2^{4-1}_{IV} ±D=ABC	2^{5-2}_{III} ±D=AB ±E=AC	2^{6-3}_{III} ±D=AB ±E=AC ±F=BC	2^{7-4}_{III} ±D=AB ±E=AC ±F=BC ±G=ABC		
16	2^{3} *twice*	2^{4} *full*	2^{5-1}_{V} ±E=ABCD	2^{6-2}_{IV} ±E=ABC ±F=ABD	2^{7-3}_{IV} ±E=ABC ±F=ABD ±G=ACD	2^{8-4}_{IV} ±E=ABC ±F=ABD ±G=ACD ±H=BCD	2^{9-5}_{III}
32	2^{3} *4 times*	2^{4} *twice*	2^{5} *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABC ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=ACDE	2^{9-4}_{IV}
64	2^{3} *8 times*	2^{4} *4 times*	2^{5} *twice*	2^{6} *full*	2^{7-1}_{VII} ±G=ABCDEF	2^{8-2}_{V} ±G=ABCD ±H=ABEF	2^{9-3}_{IV}

increasing information about additional factors

lower resolution greater aliasing

What the design's resolution tells us

Resolution III designs

- Excellent for an initial screening
  - e.g. developing a new product
  - e.g. troubleshooting a process, such as moving production from one location to another, but struggling to get a similar product on the two different machines

Resolution IV designs

- Used for learning about, and understanding, a system (characterization)

Resolution V designs and higher, and full factorial designs

- Used for optimizing a process, understanding complex effects
- To develop high-accuracy models
- You will need to justify the budget for this carefully: expensive*!*

5.9.5 Highly fractionated designs

Running a half-fraction of a 2^k factorial is not the only way to reduce the number of runs. In general, we can run a 2^{k-p} fractional factorial. A system with 2^{k-1} is called a *half fraction*, while a 2^{k-2} design is a quarter fraction.

The purpose of a fractionated design is to reduce the number of experiments when your budget or time does not allow you to complete a full factorial. Also, the full factorial is often not required, especially when k is greater than about 4, since the higher-order interaction terms are almost always insignificant. If we have a budget for 8 experiments, then we could run a:

- 2^3 full factorial on **3 factors**
- 2^{4-1} half fraction, investigating **4 factors**
- 2^{5-2} quarter fraction looking at the effects of **5 factors**
- 2^{6-3} fractional factorial with **6 factors**, or a
- 2^{7-4} fractional factorial with **7 factors**.

At the early stages of our work we might prefer to screen many factors, k, accepting a very complex confounding pattern, because we are uncertain which factors actually affect our response. Later, as we are optimizing our process, particularly as we approach an optimum, then the 2 factor and perhaps 3-factor interactions are more dominant. So investigating and calculating these effects more accurately and more precisely becomes important and we have to use full factorials. But by then we have hopefully identified much fewer factors k than what we started off with.

So this section is concerned with the trade-offs as we go from a full factorial with 2^k runs to a highly fractionated factorial, 2^{k-p}.

Example 1

You have identified 7 factors that affect your response.

3. So the 4 generators we used are **I = ABD**, **I = ACE**, **I = BCF** and **I = ABCG**. The generators such as **ABD** and **ACE** are called *"words"*.

4. The *defining relationship* is a sequence of words which are all equal to **I**. The defining relation is found from the product of all possible generator combinations, and then simplified to be written as **I = ...**

   The rule is that a 2^{k-p} factorial design is produced by p generators and has a defining relationship of 2^p words. So in this example there are $2^4 = 16$ words in our defining relation. They are:

   – Intercept: **I** [1]

   – Generators: **I = ABD = ACE = BCF = ABCG** [2,3,4,5]

   – Two combinations of generators: **I = BDCE = ACDF = CDG = ABEF = BEG = AFG** [6 to 11]

   – Three combinations of generators: **I = DEF = ADEG = CEFG = BDFG** [12 to 15]

   – Four combinations of generators: **I = ABCDEFG** [16]

   The 16 words in the defining relationship are written as: **I = ABD = ACE = BCF = ABCG = BCDE = ACDF = CDG = ABEF = BEG = AFG = DEF = ADEG = CEFG = BDFG = ABCDEFG**. The shortest length word has 3 letters.

General approach for figuring out the aliasing *before starting* the experiments

Define the number of factors to investigate; and determine what your budget is.

1. Read the generators from the trade off table
2. Rearrange the generators as $\mathbf{I} = \ldots$
3. Form the defining relationship taking all combinations of the words, so that $\mathbf{I} = \ldots$
4. Ensure the defining relationship has 2^p words
5. Compute the aliasing pattern
6. Is the aliasing problematic?

   If **yes**: reassign factor letters; or pick another design (start over)

   If **no**: you are ready to start your experiments

[Flickr: ajc1]

increasing cost

Number of factors, *k*

Number of runs	3	4	5	6	7	8	9
4	2^{3-1}_{III} ±C=AB						
8	2^{3} *full*	2^{4-1}_{IV} ±D=ABC	2^{5-2}_{III} ±D=AB ±E=AC	2^{6-3}_{III} ±D=AB ±E=AC ±F=BC	2^{7-4}_{III} ±D=AB ±E=AC ±F=BC ±G=ABC		
16	2^{3} *twice*	2^{4} *full*	2^{5-1}_{V} ±E=ABCD	2^{6-2}_{IV} ±E=ABC ±F=ABD	2^{7-3}_{IV} ±E=ABC ±F=ABD ±G=ACD	2^{8-4}_{IV} ±E=ABC ±F=ABD ±G=ACD ±H=BCD	2^{9-5}_{III}
32	2^{3} *4 times*	2^{4} *twice*	2^{5} *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABC ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=ACDE	2^{9-4}_{IV}
64	2^{3} *8 times*	2^{4} *4 times*	2^{5} *twice*	2^{6} *full*	2^{7-1}_{VII} ±G=ABCDEF	2^{8-2}_{V} ±G=ABCD ±H=ABEF	2^{9-3}_{IV}

increasing information about additional factors

lower resolution greater aliasing

ADVICE

pick a design that meets the objective

- If you are just starting out, avoid eliminating factors to simply get a full factorial.
- Use the experimental evidence to eliminate factors.

- Remember: these are experimental building blocks. The experiments you run first can be extended on later.

- In the next example, we show how factors are eliminated, *based on evidence*.

[Flickr: rahego]