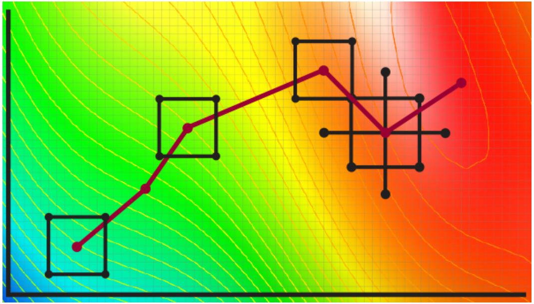
Experimentation for Improvement



http://learnche.org/

Design and Analysis of Experiments

Copyright, sharing, and attribution notice

There's more: we never stop learning!

There is ongoing research on how to efficiently run fewer experiments.

We give some pointers to this research at the end of this module.

In this section we only cover established techniques.

Cell-culture example: long duration runs; and many factors are possible

1. **T**: the temperature profile
2. **D**: dissolved oxygen
3. **A**: agitation rate
4. **P**: pH
5. **S**: substrate type (A or B)

Industrial scale:

[Flickr: londonmatt]

At 10 days per cell culture, it would take ≈ 1 year for all $2^5 = 32$ runs

Cell-culture example: long duration runs; and many factors are possible

Laboratory scale:

1. **T**: the temperature profile
2. **D**: dissolved oxygen
3. **A**: agitation rate
4. **P**: pH
5. **S**: substrate type (A or B)

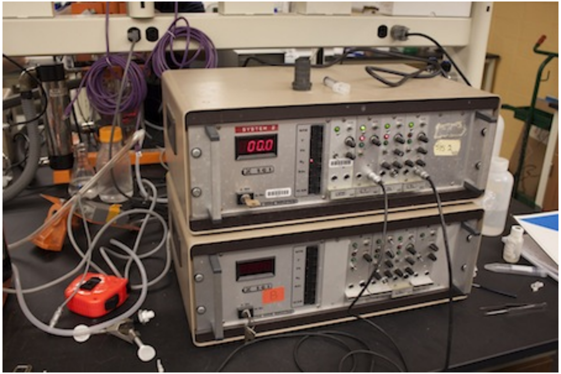
[Flickr: kaibara]

At 10 days per cell culture, it would take ≈ 1 year for all $2^5 = 32$ runs

Cell-culture example: long duration runs; and many factors are possible

1. **T**: the temperature profile
2. **D**: dissolved oxygen
3. **A**: agitation rate
4. **P**: pH
5. **S**: substrate type (A or B)

Laboratory equipment to control the culture:

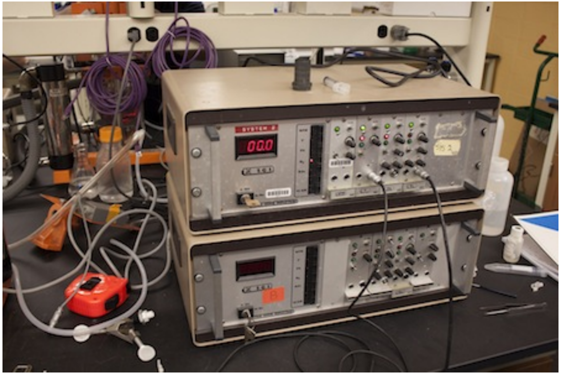


[Flickr: mjanicki]

3 months available: that corresponds to 9 experiments.

Cell-culture example: long duration runs; and many factors are possible

Laboratory equipment to control the culture:



[Flickr: mjanicki]

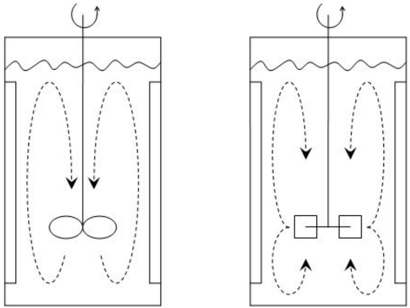
1. **T**: the temperature profile
2. **D**: dissolved oxygen
3. ~~**A**: agitation rate~~
4. ~~**P**: pH~~
5. **S**: substrate type (A or B)

Don't do this: remove factors in order to get a full factorial.

Cell-culture example: long duration runs; and many factors are possible

Different types of mixers (impellers):

1. **T**: the temperature profile
2. **D**: dissolved oxygen
3. **A**: agitation rate
4. **P**: pH
5. **S**: substrate type (A or B)
6. **W**: water type (distilled or tap water)
7. **M**: mixer type (axial or radial)



[Wikipedia]

3 months available: that corresponds to 9 experiments.

Number of factors, k

increasing cost

Number of runs	3	4	5	6	7	8
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=BCD	2^{7-3}_{IV} ±E=ABC ±F=BCD ±G=ACD	2^{8-4}_{IV} ±E=BCD ±F=ACD ±G=ABC ±H=ABD
32	2^{3} *4 times*	2^{4} *twice*	2^{5} *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABCD ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=BCDE
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

more factors, lower resolution

fewer factors, higher resolution

$$2^{5-2}$$

$$\mathbf{D} = \mathbf{AB}$$

$$\mathbf{E} = \mathbf{AC}$$

Number of factors, k

Number of runs	3	4	5	6	7	8
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=BCD	2^{7-3}_{IV} ±E=ABC ±F=BCD ±G=ACD	2^{8-4}_{IV} ±E=BCD ±F=ACD ±G=ABC ±H=ABD
32	2^3 *4 times*	2^4 *twice*	2^5 *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABCD ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=BCDE
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

increasing cost ↓

more factors, lower resolution →

← fewer factors, higher resolution

$$2^{k-p}$$

$$\mathbf{D} = \mathbf{AB}$$

$$\mathbf{E} = \mathbf{AC}$$

Number of factors, *k*

increasing cost

Number of runs	3	4	5	6	7	8
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=BCD	2^{7-3}_{IV} ±E=ABC ±F=BCD ±G=ACD	2^{8-4}_{IV} ±E=BCD ±F=ACD ±G=ABC ±H=ABD
32	2^{3} *4 times*	2^{4} *twice*	2^{5} *full*	2^{6-1}_{VI} ±F=ABCDE	2^{7-2}_{IV} ±F=ABCD ±G=ABDE	2^{8-3}_{IV} ±F=ABC ±G=ABD ±H=BCDE
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

more factors, lower resolution

fewer factors, higher resolution

$$2^{k-p}$$

$$p = 1$$

Number of factors, *k*

increasing cost

Number of runs

Number of runs	3	4	5	6	7	8
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=BCD**	2^{7-3}_{IV} **±E=ABC** **±F=BCD** **±G=ACD**	2^{8-4}_{IV} **±E=BCD** **±F=ACD** **±G=ABC** **±H=ABD**
32	2^{3} *4 times*	2^{4} *twice*	2^{5} *full*	2^{6-1}_{VI} **±F=ABCDE**	2^{7-2}_{IV} **±F=ABCD** **±G=ABDE**	2^{8-3}_{IV} **±F=ABC** **±G=ABD** **±H=BCDE**
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**

$$2^{k-p}$$

$$p = 2$$

more factors, lower resolution

fewer factors, higher resolution

Cell-culture example: creating the fractional factorial design

Experiment	$\mathbf{A}^*$	$\mathbf{B}^\circ$	$\mathbf{C}^\circ$	$\mathbf{D}^\circ = \mathbf{AB}$	$\mathbf{E}^* = \mathbf{AC}$	
1	$-$	$-$	$-$	$+$	$+$	
2	$+$	$-$	$-$	$-$	$-$	
3	$-$	$+$	$-$	$-$	$+$	
4	$+$	$+$	$-$	$+$	$-$	
5	$-$	$-$	$+$	$+$	$-$	
6	$+$	$-$	$+$	$-$	$+$	
7	$-$	$+$	$+$	$-$	$-$	
8	$+$	$+$	$+$	$+$	$+$	
9	-1	0	0	0	$+1$	$\longleftarrow$ this row is a baseline #

*categorical factor

$^\circ$ continuous factor

#this entry does **not** use the generators; it should be run first to establish a baseline

Aliasing: when we have more than one name for the same thing

What is aliased in this experimental design (i.e. which columns are the same)?

- **A=BC**
- **B=AC**
- **C=AB**
- **ABC = Intercept** (the intercept is indicated as b_0)

Calculating the aliases using an interesting technique

$2^{5\text{-}2}_{\mathrm{III}}$

±D=AB

±E=AC

$$\mathbf{D} = \mathbf{AB}$$

$$\mathbf{D\,D} = \mathbf{AB\,D}$$

rule: **AA=I**; **BB=I**; ... **DD=I**, *etc*

$$\mathbf{I} = \mathbf{ABD}$$

If we have $\mathbf{A} = \begin{pmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \end{pmatrix}$ then we can say $\mathbf{AA} = \begin{pmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \end{pmatrix} \begin{pmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \end{pmatrix} = \begin{pmatrix} +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \end{pmatrix} = \mathbf{I}$

Calculating the aliases using an interesting technique

2^{5-2}_{III}

$\pm D = AB$

$\pm E = AC$

$$\mathbf{E} = \mathbf{AC}$$

$$\mathbf{E\,E} = \mathbf{AC\,E}$$

rule: $\mathbf{AA{=}I}$; $\mathbf{BB{=}I}$; ... $\mathbf{DD{=}I}$, *etc*

$$\mathbf{I} = \mathbf{ACE}$$

Multiple the left and right by the symbol that's on the left.

Approach to calculate the aliasing pattern

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

1. **D = AB** and **E = AC**
2. **I = ABD** and **I = ACE**
3. **I = ABD = ACE = (ABD)(ACE)**
4. $p = 2$, and we have 4 words.
5. See the next slides.

Word: a collection of sequential factor letters

(ABD)(ACE) = AABCDE = (AA)BCDE = (I)BCDE = BCDE

Approach to calculate the aliasing pattern

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

1. **D = AB** and **E = AC**
2. **I = ABD** and **I = ACE**
3. **I = ABD = ACE = BCDE**
4. $p = 2$, and we have 4 words.
5. We will in the next video.

Word: a collection of sequential factor letters

(ABD)(ACE) = AABCDE = (AA)BCDE = (I)BCDE = BCDE

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. In the next video: use this defining relationship to compute the aliasing pattern

Example to try yourself: 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 next.

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

±E=ABC

±F=BCD