Factorial Designs

BIOE 498/598 PJ

Spring 2021

What is a factorial design?

A factorial design studies multiple factors set at discrete intervals.

What is a factorial design?

A factorial design studies multiple factors set at discrete intervals.

A design with k factors set at L levels is called an L^k factorial design.

What is a factorial design?

A factorial design studies multiple factors set at discrete intervals.

A design with k factors set at L levels is called an L^k factorial design.

A factorial design includes runs with every combination of factors set at every level.

Sample factorial designs

2^2 Factorial design

x_1	x_2
$-$	$-$
$+$	$-$
$-$	$+$
$+$	$+$

2^3 Factorial design

x_1	x_2	x_3
$-$	$-$	$-$
$+$	$-$	$-$
$-$	$+$	$-$
$+$	$+$	$-$
$-$	$-$	$+$
$+$	$-$	$+$
$-$	$+$	$+$
$+$	$+$	$+$

3^2 Factorial design

x_1	x_2
$-$	$-$
0	$-$
$+$	$-$
$-$	0
0	0
$+$	0
$-$	$+$
0	$+$
$+$	$+$

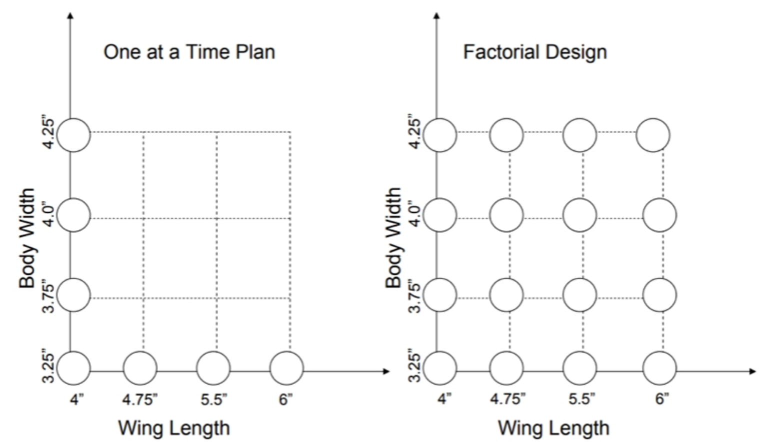
Why do we use factorial designs?

- Factorial designs find better optima.
- Factorial designs are more efficient.
- Factorial designs make better estimates of effect sizes.

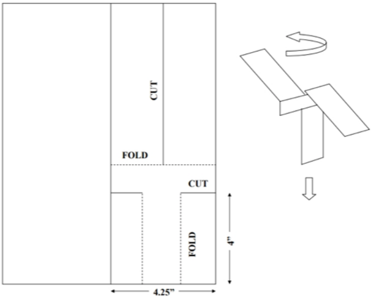
Why do we use factorial designs?

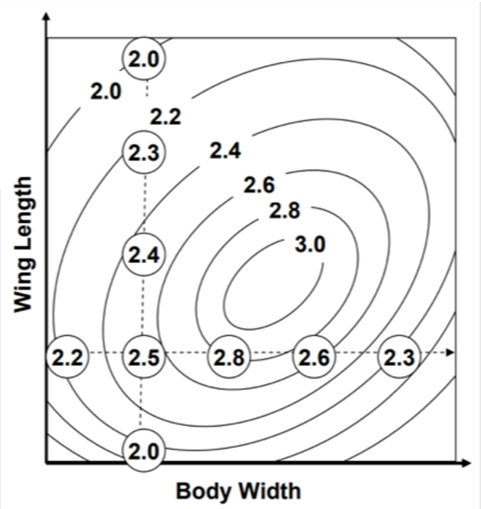
- **Factorial designs find better optima.**
- Factorial designs are more efficient.
- Factorial designs make better estimates of effect sizes.

What is a factorial design?

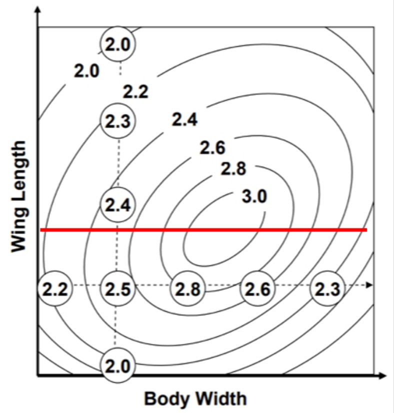


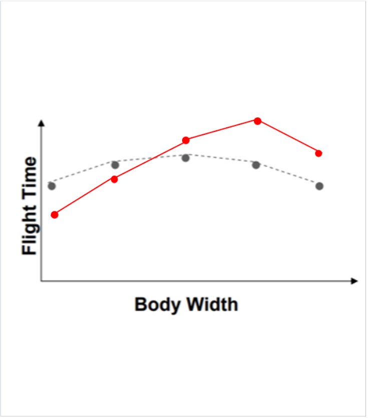
Factorial designs find better optima





The problem with OFAT: Interactions



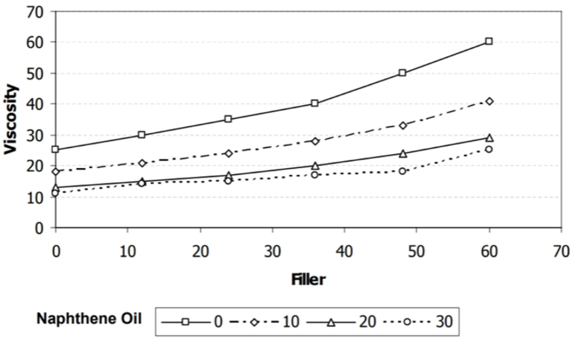


Using interaction plots for diagnosis

Table 3.1 *Mooney Viscosity of Silica B at* $100^\circ C$

Naphthene Oil (phr)	Filler (phr)					
	0	12	24	36	48	60
0	25	30	35	40	50	60
10	18	21	24	28	33	41
20	13	15	17	20	24	29
30	11	14	15	17	18	25

Figure 3.4 *Interaction Plot of Filler and Naphthene Oil*



Why do we use factorial designs?

- Factorial designs find better optima.
- **Factorial designs are more efficient.**
- Factorial designs make better estimates of effect sizes.

Factorial designs seem *less* efficient...

Imagine an experiment with four factors, each with two levels $(-, +)$.
We want three replicates for each level.

One Factor at a Time Design

- 3 runs at level $(-)$
- 4 factors $\times$ 3 runs at $(+) = 12$ runs
- **15 total runs**

Factorial Design

- $2^4 =$ **16 total runs**

. . . until you look at the designs

OFAT design

x_1	x_2	x_3	x_4
−	−	−	−
−	−	−	−
−	−	−	−
+	−	−	−
+	−	−	−
+	−	−	−
−	+	−	−
−	+	−	−
−	+	−	−
−	−	+	−
−	−	+	−
−	−	+	−
−	−	−	+
−	−	−	+
−	−	−	+

Factorial design

x_1	x_2	x_3	x_4
−	−	−	−
+	−	−	−
−	+	−	−
+	+	−	−
−	−	+	−
+	−	+	−
−	+	+	−
+	+	+	−
−	−	−	+
+	−	−	+
−	+	−	+
+	+	−	+
−	−	+	+
+	−	+	+
−	+	+	+
+	+	+	+

Factorial designs give more replicates per run

A factorial design in n variables has 2^n runs, but 2^{n-1} replicates at each level $(-, +)$.

Factorial designs give more replicates per run

A factorial design in n variables has 2^n runs, but 2^{n-1} replicates at each level $(-, +)$.

Imagine a design with n variables at k levels.
After the initial design, adding another replicate requires

- nk runs for a OFAT design
- $\sim k$ runs for a factorial design

Why do we use factorial designs?

- Factorial designs find better optima.
- Factorial designs are more efficient.
- **Factorial designs make better estimates of effect sizes.**

What are the other factors doing when x_3 is high?

OFAT design

x_1	x_2	x_3	x_4
−	−	−	−
−	−	−	−
−	−	−	−
+	−	−	−
+	−	−	−
+	−	−	−
−	+	−	−
−	+	−	−
−	+	−	−
−	−	+	−
−	−	+	−
−	−	+	−
−	−	−	+
−	−	−	+
−	−	−	+

Factorial design

x_1	x_2	x_3	x_4
−	−	−	−
+	−	−	−
−	+	−	−
+	+	−	−
−	−	+	−
+	−	+	−
−	+	+	−
+	+	+	−
−	−	−	+
+	−	−	+
−	+	−	+
+	+	−	+
−	−	+	+
+	−	+	+
−	+	+	+
+	+	+	+

What do the effect sizes estimate?

For OFAT designs:

β_i is the effect of moving x_i from $-$ to $+$
while all other factors stay at $-$.

What do the effect sizes estimate?

For OFAT designs:

β_i is the effect of moving x_i from $-$ to $+$
while all other factors stay at $-$.

For factorial designs:

β_i is the effect of moving x_i from $-$ to $+$
averaged over all other factors at all levels.

Factorial designs are nested

Factorial design

x_1	x_2	x_3	x_4
$-$	$-$	$-$	$-$
$+$	$-$	$-$	$-$
$-$	$+$	$-$	$-$
$+$	$+$	$-$	$-$
$-$	$-$	$+$	$-$
$+$	$-$	$+$	$-$
$-$	$+$	$+$	$-$
$+$	$+$	$+$	$-$
$-$	$-$	$-$	$+$
$+$	$-$	$-$	$+$
$-$	$+$	$-$	$+$
$+$	$+$	$-$	$+$
$-$	$-$	$+$	$+$
$+$	$-$	$+$	$+$
$-$	$+$	$+$	$+$
$+$	$+$	$+$	$+$

When $x_3 = -$

x_1	x_2	x_4
$-$	$-$	$-$
$+$	$-$	$-$
$-$	$+$	$-$
$+$	$+$	$-$
$-$	$-$	$+$
$+$	$-$	$+$
$-$	$+$	$+$
$+$	$+$	$+$

When $x_3 = +$

x_1	x_2	x_4
$-$	$-$	$-$
$+$	$-$	$-$
$-$	$+$	$-$
$+$	$+$	$-$
$-$	$-$	$+$
$+$	$-$	$+$
$-$	$+$	$+$
$+$	$+$	$+$

Rank revisited

The rank of a matrix quantifies the number of linearly independent rows or columns.

The column rank of a matrix is always equal to the row rank.

$$\mathrm{rank}(\mathsf{X}) = \mathrm{rank}(\mathsf{X}^{\mathrm{T}})$$

This limits the rank to be at most the smaller dimension of the matrix.

$$\mathrm{rank}(\mathsf{X}) \leq \min\{m, n\} \quad \text{if} \quad \dim(\mathsf{A}) = m \times n$$

If the above *equality* holds, we say that the matrix is **full rank**.

Rank and linear modeling

Each parameter in a linear model requires one independent piece of information.

The linear model $y = X\beta + \epsilon$ is solvable if and only if the design matrix X is full rank.

Solvability vs. power

A model is solvable if the design matrix is full rank. However, we need "extra" rows (degrees of freedom) to estimate the model's uncertainty.

Consider the one parameter model $y = \beta x + \epsilon$.
Given data $(x,y) = (3,6)$:

$$\hat{\beta} = x^{-1} y = 2$$

Solvability vs. power

A model is solvable if the design matrix is full rank. However, we need "extra" rows (degrees of freedom) to estimate the model's uncertainty.

Consider the one parameter model $y = \beta x + \epsilon$.
Given data $(x,y) = (3,6)$:

$$\hat{\beta} = x^{-1} y = 2$$

Substituting back we see that

$$\epsilon = y - \hat{\beta} x = 6 - 2 \times 3 = 0$$

With one data point our estimate is always exact!

Solvability vs. power

A model is solvable if the design matrix is full rank. However, we need "extra" rows (degrees of freedom) to estimate the model's uncertainty.

Consider the one parameter model $y = \beta x + \epsilon$.
Given data $(x,y) = (3,6)$:

$$\hat{\beta} = x^{-1} y = 2$$

Substituting back we see that

$$\epsilon = y - \hat{\beta} x = 6 - 2 \times 3 = 0$$

With one data point our estimate is always exact!
Now let's use two data points: $(x,y) = (3,6)$ and $(x,y) = (4,12)$.

$$\hat{\beta} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}^{+} \begin{pmatrix} 6 \\ 12 \end{pmatrix} = 2.64$$

$$\epsilon_1 = y_1 - \hat{\beta} x = 6 - 2.64 \times 3 = -1.92$$

$$\epsilon_2 = y_2 - \hat{\beta} x = 12 - 2.64 \times 4 = 1.44$$

What does this mean for factorial designs?

A full factorial design with n variables has 2^n experiments. It also has 2^n coefficients (intercept, first-order, and interaction). We can fit a model to a full factorial design but will have no information leftover to estimate the error.

We have three options if we want statistical power behind our factorial designs:

1. perform replicates of some (or all) runs
2. only estimate a subset of the 2^n coefficients
3. some combination of 1 & 2

What does this mean for factorial designs?

A full factorial design with n variables has 2^n experiments. It also has 2^n coefficients (intercept, first-order, and interaction). We can fit a model to a full factorial design but will have no information leftover to estimate the error.

We have three options if we want statistical power behind our factorial designs:

1. perform replicates of some (or all) runs
2. only estimate a subset of the 2^n coefficients
3. some combination of 1 & 2

- For small n designs we perform replicates since there are already few runs and the interactions are probably significant.
- For large n designs we drop coefficients for higher order terms since we already have lots of runs and the higher-order interactions are most likely zero.

Summary

Factorial designs produce more information with fewer runs than OFAT designs.

- Factorial designs find better optima.
- Factorial designs are more efficient.
- Factorial designs make better estimates of effect sizes.

Summary

Factorial designs produce more information with fewer runs than OFAT designs.

- Factorial designs find better optima.
- Factorial designs are more efficient.
- Factorial designs make better estimates of effect sizes.

As the number of factors increases, factorial designs become **more** efficient.

However, the number of runs can be prohibitive, and we rarely need to estimate higher-order interactions.

Summary

Factorial designs produce more information with fewer runs than OFAT designs.

- Factorial designs find better optima.
- Factorial designs are more efficient.
- Factorial designs make better estimates of effect sizes.

As the number of factors increases, factorial designs become **more** efficient.

However, the number of runs can be prohibitive, and we rarely need to estimate higher-order interactions.

Fractional factorial designs require fewer runs by purposefully ignoring the higher-order terms.