Screening Designs

BIOE 498/598 PJ

Spring 2022

Why do we use screening designs?

- Optimization is expensive—many runs/factor at > 2 levels
- Too many factors waste resources
- Too few factors lead to suboptimal results

Why do we use screening designs?

- Optimization is expensive—many runs/factor at > 2 levels
- Too many factors waste resources
- Too few factors lead to suboptimal results

- **Solution:** A *screening design* tests a large number of factors
- Only active factors are carried forward for optimization

What is a screening design?

- Screening designs have few runs, ideally ≤ 2 runs/factor.

What is a screening design?

- Screening designs have few runs, ideally ≤ 2 runs/factor.
- The focus is on main effects. By the Effect Hierarchy and Effect Heredity principles,

$$\text{important factor} \approx \text{significant main effect}$$

What is a screening design?

- Screening designs have few runs, ideally ≤ 2 runs/factor.
- The focus is on main effects. By the Effect Hierarchy and Effect Heredity principles,

$$\text{important factor} \approx \text{significant main effect}$$

- We don't worry about estimates of TWIs. We're selecting factors, not interactions.

Types of screening designs

- Resolution III Fractional Factorial Design
  - Pro: Mirror image can clear main effects
  - Con: Run size always a power of 2

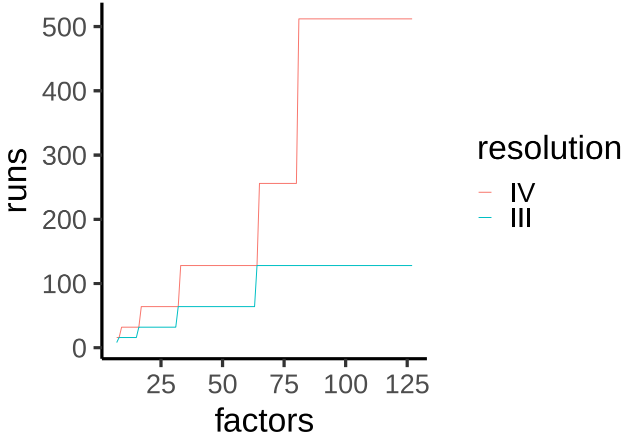
Types of screening designs

- Resolution III Fractional Factorial Design
  - Pro: Mirror image can clear main effects
  - Con: Run size always a power of 2
- PB Design
  - Pro: Run size in multiples of 4
  - Con: Complex aliasing

Types of screening designs

- Resolution III Fractional Factorial Design
  - Pro: Mirror image can clear main effects
  - Con: Run size always a power of 2
- PB Design
  - Pro: Run size in multiples of 4
  - Con: Complex aliasing
- Definitive Screening Designs
  - Hybrid screening/optimization design. We'll discuss later!

Don't rule out Fractional Factorial Designs.



number of runs

number of factors	8	16	32	64	128	256	512	1024	2048	4096
						only the MA design				
3	full									
4	IV	full								
5	III	V	full							
6	III	IV	VI	full						
7	III	IV	IV	VII	full					
8		IV	IV	V	VIII	full				
9		III	IV	IV	VI	IX	full			
10		III	IV	IV	V	VI	X	full		
11		III	IV	IV	V	VI	VII	XI	full	
12		III	IV	IV	IV	VI	VI	VIII	XII	full
13		III	IV	IV	IV	V	VI	VII	VIII	XIII
14		III	IV	IV	IV	V	VI	VII	VIII	IX
15		III	IV	IV	IV	V	VI	VII	VIII	VIII
16			IV	IV	IV	V	VI	VI	VIII	VIII
17			III	IV	IV	V	VI	VI	VII	VIII
18			III	IV	IV	IV	VI	VI	VII	VIII
19			III	IV	IV	IV	V	VI	VII	VIII
20			III	IV	IV	IV	V	VI	VII	VIII
21			III	IV	IV	IV	V	VI	VII	VIII
22			III	IV	IV	IV	V	VI	VII	VIII
23			III	IV	IV	IV	V	VI	VII	VIII
24			III	IV	IV	IV	IV	VI	VI	VIII

	32	64	128	256	512	1024	2048	4096
Resolution III up to	31	63	127			factors.		
Resolution IV up to		32	64	80	160	factors.		
Resolution V up to number of factors:						33	47	65
Resolution VI up to number of factors:						24	34	48
First design is MA up to number of factors:	31	63	127	36	29	28	32	26

Gromping, 2014
J. Stat. Software

Workflow for Resolution III screens

1. Run the design
2. Fit the model with main effects. If you have DoF left over, add any TWIs that are **not** confounded with main effects.
3. If the overall model fit is bad, or if you expected certain effects to be significant that were not, consider a second batch of runs with a mirror image design.
4. Drop any factors that are not **important** (practically or statistically).

Plackett-Burman Designs

- Discovered in 1946 while working in the British Ministry of Supply
- Orthogonal designs, so main effects can be estimated independently
- Run sizes in **multiples of 4**

Plackett-Burman Designs

- Discovered in 1946 while working in the British Ministry of Supply
- Orthogonal designs, so main effects can be estimated independently
- Run sizes in **multiples of 4**

- Both PB designs and FF designs are *Orthogonal Arrays*
  - PB = FF when $N = 2^k$

Plackett-Burman Designs

- Discovered in 1946 while working in the British Ministry of Supply
- Orthogonal designs, so main effects can be estimated independently
- Run sizes in **multiples of 4**
- Both PB designs and FF designs are *Orthogonal Arrays*
  - PB = FF when $N = 2^k$
- PB designs have *complex aliasing*. Every ME is partially confounded with all TWIs.

Creating a PB design (up to 23 factors)

1. Start with the first run from the following table.

Runs	Factor Levels
12	$++-+++---+-$
20	$++--++++-+-+----++-$
24	$+++++-+-++--++--+-+----$

2. Cycle the factor levels by one to get run #2. Repeat for 11, 19, or 23 runs.
3. Set the final run to all low ($-$).
4. If the number of factors k is less than the number of runs, select the first k columns.

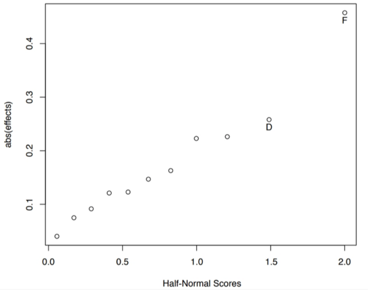
Workflow for PB designs

1. Run the design.
2. Fit a model with main effects plus an effect for any unused column in the design.
3. Optional: Perform subset regression to identify factors that appear frequently in smaller models with good predictive power.
4. Drop any factors that are not **important** (practically or statistically).
5. If only a small number of factors remain, try refitting the small model.

Example PB design: Cast fatigue

Table 6.11 *Design Matrix and Lifetime Data for Cast Fatigue Experiment*

Run	A	B	C	D	E	F	G	c8	c9	c10	c11	
1	+	−	+	+	+	−	−	−	+	−	+	4.733
2	−	+	+	+	−	−	−	+	−	+	+	4.625
3	+	+	+	−	−	−	+	−	+	+	−	5.899
4	+	+	−	−	−	+	−	+	+	−	+	7.000
5	+	−	−	−	+	−	+	+	−	+	+	5.752
6	−	−	−	+	−	+	+	−	+	+	+	5.682
7	−	−	+	−	+	+	−	+	+	+	−	6.607
8	−	+	−	+	+	−	+	+	+	−	−	5.818
9	+	−	+	+	−	+	+	+	−	−	−	5.917
10	−	+	+	−	+	+	+	−	−	−	+	5.863
11	+	+	−	+	+	+	−	+	−	+	−	6.058
12	−	−	−	−	−	−	−	−	−	−	−	4.809



This design includes 7 factors; however, effects are estimated for all columns. The last 4 "factors" are interactions with complex aliasing.

To replicate or not to replicate?

- Many screening designs are *saturated* — there are no DoF to estimate confidence intervals for the parameters.
- The number of estimable factors is bounded by the rank of the model matrix. Replicates do not changes the rank.
- If you don't replicate the design, you can select factors based on the magnitude of the effects alone (half-normal plot).
  - Remember that half-normal plots work better as the number of factors grows.

To replicate or not to replicate?

- Many screening designs are *saturated* — there are no DoF to estimate confidence intervals for the parameters.
- The number of estimable factors is bounded by the rank of the model matrix. Replicates do not changes the rank.
- If you don't replicate the design, you can select factors based on the magnitude of the effects alone (half-normal plot).
  - Remember that half-normal plots work better as the number of factors grows.

- Replicating a Resolution III Design
  - Consider a mirror-image instead. This will give clear main effects.
  - Check if you can afford a Resolution IV instead. This gives clear main effects and a confounding structure.
- Replicating a PB Design
  - Replicating the design will help you estimate the "pure error".
  - You can "move up" to a larger PB design to get extra runs. This won't estimate pure error, but you can add more confounded effects to the model to improve the estimates.