Using Optimal Designs to Solve Practical Experimental Problems

Pat Whitcomb, Founder Stat-Ease, Inc.
pat@statease.com

Copyright 2020 Stat-Ease, Inc.
September 2020
Making the most of this learning opportunity

To prevent audio disruptions, all attendees will be muted.

Questions can be posted in the Question area. If they are not addressed during the webinar, I will reply via email afterwards.

Questions may also be sent to stathelp@statease.com. Please provide your company name and, if you are using Design-Expert, the serial number (found under Help, About).

Note: The slides and a recording of this webinar will be posted on the Webinars page of the Stat-Ease website within a few days.
Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- Optimal point selection ($I$ versus $D$ optimality).
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.
Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- Optimal point selection ($I$ versus $D$ optimality).
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.
Study Considerations

1. What is the objective of the study?
2. State the objective in terms of measured responses:
   - How will the responses be measured?
   - What precision is required?
3. Which factors will be studied?
4. What are the regions of interest and operability?
5. What order polynomial will adequately model response behavior?
6. What design should we use?
“Good” Response Surface Designs
A Checklist

1. Allow the polynomial chosen by the experimenter to be estimated well.

2. Give sufficient information to allow a test for lack of fit.
   - Have more unique design points than coefficients in model.
   - Provide an estimate of “pure” error.

3. Be insensitive (robust) to the presence of outliers in the data.

4. Be robust to errors in control of the factor levels.

5. Permit blocking and sequential experimentation.

6. Provide a check on homogeneous variance assumption and other useful model diagnostics; including deletion statistics.

7. Generate useful information throughout the region of interest, i.e., provide a good distribution of standard error of prediction.

8. Not contain an excessively large number of runs.

Copyright 2020 Stat-Ease, Inc.
“designing an experiment is not necessarily easy and should involve balancing multiple objectives, not just focusing on single characteristic.”


“Alphabetic optimality is not enough!”

Pat speaking for Stat-Ease, Inc.
Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- **Optimal point selection (I versus D optimality).**
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.
Goal: D-optimal design minimizes the determinant of the $(X'X)^{-1}$ matrix. This minimizes the volume of the confidence ellipsoid for the coefficients and maximizes information about the polynomial coefficients.
An **I-optimal** design seeks to minimize the integral of the prediction variance across the design space. These designs are built algorithmically to provide lower integrated prediction variance across the design space. This equates to minimizing the area under the FDS curve.
Optimal Point Selection
I versus D Optimal Design

Compare point selection using I-optimal and D-optimal:

- Build a one factor design.
- Design for a quadratic model.
- Choose all twelve runs using optimality as only criterion.
I versus D Optimal Design
Optimal 12 Point Designs

Using Optimal Designs to Solve Practical Experimental Problems
I-optimal versus D-optimal
One Factor 12 Optimal Points

Fraction of Design Space Graph

I min: 0.382
I avg: 0.421
I max: 0.577
D min: 0.395
D avg: 0.447
D max: 0.500

Copyright 2020 Stat-Ease, Inc.

Using Optimal Designs to Solve Practical Experimental Problems
What about G-Optimality?
Three 6-point 2-factor Designs

<table>
<thead>
<tr>
<th></th>
<th>G-optimal</th>
<th>D-optimal</th>
<th>I-optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>G efficiency</td>
<td>87.9%</td>
<td>66.4%</td>
<td>56.5%</td>
</tr>
<tr>
<td>Min SE mean</td>
<td>0.653</td>
<td>0.604</td>
<td>0.566</td>
</tr>
<tr>
<td>Ave SE mean</td>
<td>0.777</td>
<td>0.743</td>
<td>0.699</td>
</tr>
<tr>
<td>Max SE mean</td>
<td>0.923</td>
<td>1.063</td>
<td>1.152</td>
</tr>
</tbody>
</table>

G-optimal designs:
- Minimize the maximum predicted variance.
- This is at the expense of the average prediction variance.
- For a gain in a very small fraction of the design space, precision is sacrificed in the vast majority of the design space. *(see next slide)*
What about G-Optimality?
Three 6-point 2-factor Designs

Conclusion: Do not use G!
Conclusions:

- I-optimal designs tend to place points more uniformly in the design space.
- I-optimal designs have a higher maximum prediction variance; therefore a lower G-efficiency.
- I-optimal designs have a lower average prediction variance. *(This also contributes to a lower G-efficiency.)*
- Neither design has information to evaluate sufficiency of quadratic model!
Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- Optimal point selection (I versus D optimality).
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.

Copyright 2020 Stat-Ease, Inc.
Optimal Point Selection
I versus D Optimal Design

Compare point selection using I-optimal and D-optimal:

- Build a one factor design.
- Design for a quadratic model.
- Choose eight of the twelve runs using optimality as the criteria.
- Choose four of the twelve runs as lack of fit (LOF) points using distance as the criteria.
  
  *(Maximize the minimum distance from an existing design point; i.e. fill the "holes" among the design points.)*
Using Optimal Designs to Solve Practical Experimental Problems

Optimal Designs
8 Optimal + 4 LOF Points
I-optimal versus D-optimal
One Factor 8 Optimal and 4 Distance Points

FDS Graph

I min: 0.381
I avg: 0.438
I max: 0.653
D min: 0.395
D avg: 0.448
D max: 0.547

Copyright 2020 Stat-Ease, Inc.
I-optimal versus D-optimal
12 Optimal - 8 Optimal + 4 Distance Points

FDS Graph

8 opt + 4 dist
- I min: 0.381
- I avg: 0.438
- I max: 0.653
- D min: 0.395
- D avg: 0.448
- D max: 0.547

12 optimal
- I min: 0.382
- I avg: 0.421
- I max: 0.577
- D min: 0.395
- D avg: 0.447
- D max: 0.500

Copyright 2020 Stat-Ease, Inc.

Using Optimal Designs to Solve Practical Experimental Problems
Optimal Point Selection
I versus D Optimal Design

Compare point selection for a two-factor 14-run design:

- Design for a quadratic model.

- I-optimal:
  - 14 optimal runs
  - 10 optimal and 4 LOF (*distance*)

- D-optimal:
  - 14 optimal runs
  - 10 optimal and 4 LOF (*distance*)
I-optimal Designs
14 Run Designs with 0 and 4 LOF Points

Copyright 2020 Stat-Ease, Inc.
I-optimal Designs
14 Run Designs with 0 and 4 LOF Points

FDS Graph

10 I-optimal + 4 LOF points
I min: 0.435
I avg: 0.528
I max: 0.857

14 I-optimal points
I min: 0.418
I avg: 0.515
I max: 0.908

Copyright 2020 Stat-Ease, Inc.
D-optimal Designs
14 Run Designs with 0 and 4 LOF Points

Using Optimal Designs to Solve Practical Experimental Problems

Copyright 2020 Stat-Ease, Inc.
D-optimal Designs
14 Run Designs with 0 and 4 LOF Points

FDS Graph

14 D-optimal points
Determinant of $(X'X)^{-1} = 3.906E-3$

10 D-optimal + 4 LOF points
Determinant of $(X'X)^{-1} = 5.313E-3$

Copyright 2020 Stat-Ease, Inc.

Using Optimal Designs to Solve Practical Experimental Problems
Adding LOF points:

- The design is not as alphabetically optimal.
- It is more space filling.
- Ability to detect lack of fit is enhanced.
- Adding LOF points is a good trade off!
Estimating pure error:

- In physical experiments it is desirable to build in an estimate of experimental error.
- Replicates provide an estimate of experimental error independent of model assumptions.
- Adding replicates is a good trade off!
Agenda

Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- Optimal point selection (*I* versus *D* optimality).
- Practical aspects algorithmic design.
- **Optimal design example.**
- Conclusion and recommendations.
Spray Coating Problems (Constraints)

Problems:

- At the vertex (A = 10, B = 3 and C = 0.1) not enough coating is applied.
- At the vertex (A = 30, B = 10 and C = 0.5) too much coating is applied.

<table>
<thead>
<tr>
<th>Name</th>
<th>Units</th>
<th>–1 level</th>
<th>+1 level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A flow rate</td>
<td>ml/min</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>B pressure</td>
<td>kPa</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>C linear speed</td>
<td>inch/sec</td>
<td>0.1</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Define constraint as points on edge of cuboidal space. Consider the setting for each factor that provides adequate coating while all other factors are at their low coating weight level.

- A (flow rate) $\geq 15$ when $B = 3$ and $C = 0.5$
  \[ CP_A = 15 \]

- B (pressure) $\geq 6$ when $A = 10$ and $C = 0.5$
  \[ CP_B = 6 \]

- C (linear speed) $\leq 0.3$ when $A = 10$ and $B = 3$
  \[ CP_C = 0.3 \]
Prevent “too much” Coating
(A = 30, B =10 and C = 0.1)

Define constraint as points on edge of cuboidal space. Consider the setting for each factor that provides adequate coating while all other factors are at their high coating weight level.

- A (flow rate) $\leq 20$ when B = 10 and C = 0.1
  $CP_A = 20$

- B (pressure) $\leq 6$ when A = 30 and C = 0.1
  $CP_B = 6$

- C (linear speed) $\geq 0.3$ when A = 30 and B = 10
  $CP_C = 0.3$
Prevent Not enough and Too much
Multiple Linear Constraints

**Not Enough**
exclude (10, 3, 0.5)

4.5 ≤ 0.6A + B − 15C

**Too Much**
exclude (30, 10, 0.1)

0.8A + 2B − 40C ≤ 32

Copyright 2020 Stat-Ease, Inc.

Using Optimal Designs to Solve Practical Experimental Problems
Given how many factors (k) you study and the number of coefficients (p) in the model you select, use the following as a guide to a starting design:

- **Model**: p points using an optimality criterion
- **Lack-of-Fit**: 5 points; based on distance or estimating higher order model terms.
- **Replicates**: 5 points, using the model optimality criterion (most influential).

Evaluate precision of the starting design via the FDS plot:

- If more precision is required rebuild the design adding more runs.
Spray Coating Design
20 Points: 10 I-optimal, 5 LOF, 5 replicates
Spray Coating
Evaluate your I-optimal Design

Is the optimal design precise enough?

- Want the estimated mean of the “average thickness” (thickness) to be within ± 0.5.
- The estimated standard deviation is 0.30.
Using Optimal Designs to Solve Practical Experimental Problems:

- What’s required for a good design.
- Optimal point selection (*I versus D optimality*).
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.
Top DOE choices for RSM designs:

- **Central Composite**: robust, classic design to fit quadratic model. *(Axial distances can be modified.)*

- **Box-Behnken**: good alternative 3-level design.

- **Optimal**: most flexible design. Use for:
  - designs with multiple linear constraints
  - designs with categoric or discrete numeric factors
  - models other than full quadratic
  - to augment an existing design

Always choose a design that fits the problem!

Size for precision!

Copyright 2020 Stat-Ease, Inc.
1. Identify opportunity and define objective.

2. State objective in terms of measurable responses.
   - Define the precision desired to predict each response.
   - Estimate experimental error ($\sigma$) for each response.

3. Select the input factors and ranges to study.

4. Select a design and:
   - Evaluate precision via the FDS plot.
   - Examine the design layout to ensure all the factor combinations are safe to run and are likely to result in meaningful information (no disasters).
Should I use a D-optimal or I-optimal design?

- I-optimal - precise estimation of the predictions
  Best for empirical response surface design

- D-optimal - precise estimation of model coefficients
  Best for screening and mechanistic models
Given how many factors (k) you study and the number of coefficients (p) in the model you select, use the following as a guide to a starting design:

- **Model**: p points using an optimality criteria
- **Lack-of-Fit**: 5 points; based on distance or estimating higher order model terms.
- **Replicates**: 5 points, using the model optimality criteria (most influential).

Evaluate precision of the starting design via the FDS plot:

- If more precision is required rebuild the design adding more runs.
Practical Aspects of DOE
Keep in Mind

No alphabetic optimality or sophisticated statistical analysis can make up for:

- Studying the wrong problem.
- Measuring the wrong response.
- Not having adequate precision.
- Studying the wrong factors.
- Having too many runs outside the region of operability.
References


**Stat-Ease Training:**
Sharpen Up Your DOE Skills

Modern DOE for Process Optimization
Mixture Design for Optimal Formulations

<table>
<thead>
<tr>
<th>Individuals</th>
<th>Teams (6+ people)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve your DOE skills</td>
<td>Choose your date &amp; time</td>
</tr>
<tr>
<td>Topics applicable to both novice and advanced practitioners</td>
<td>Add company case studies</td>
</tr>
</tbody>
</table>

Learn more: [www.statease.com](http://www.statease.com)

Contact: workshops@statease.com

Copyright 2020 Stat-Ease, Inc.