Stat Ease
statistics made easy
*Make the most from every experiment!*SM

# Using Optimal Designs to Solve Practical Experimental Problems

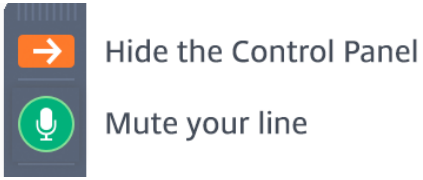
Pat Whitcomb, Founder Stat-Ease, Inc.

pat@statease.com



**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.

# 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.

# 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.

# 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.

# “Good” Response Surface Designs
## Comments on the Checklist

“designing an experiment is not necessarily easy and should involve balancing multiple objectives, not just focusing on single characteristic.”

*Response Surface Methodology*, Myers, Montgomery and Anderson-Cook, 2009, John Wiley & Sons.

“Alphabetic optimality is not enough!”

Pat speaking for Stat-Ease, Inc.

# Agenda

Using Optimal Designs to Solve Practical Experimental Problems:

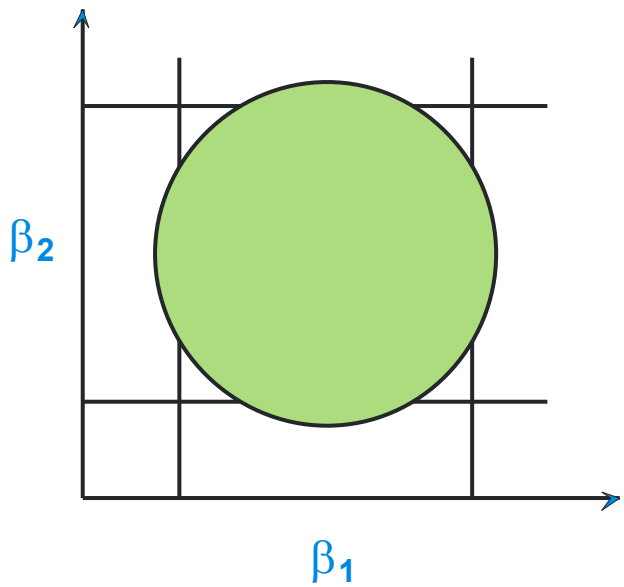
- What's required for a good design.
- **Optimal point selection (*I versus D optimality*).**
- Practical aspects algorithmic design.
- Optimal design example.
- Conclusion and recommendations.

# Optimal Point Selection

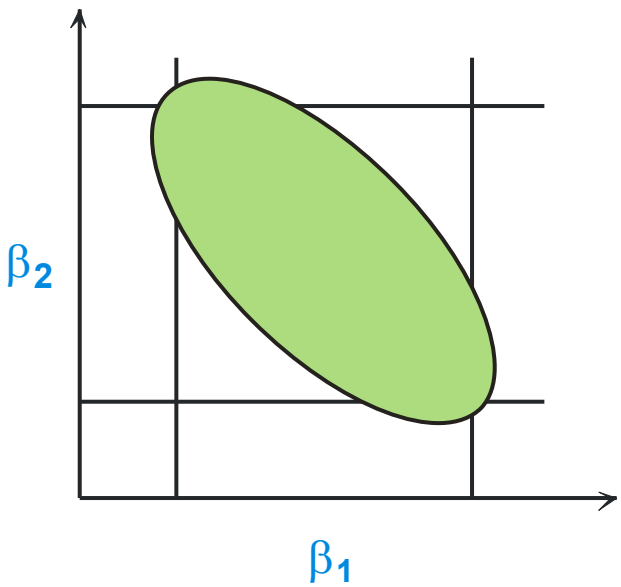
## D-optimal Point Selection

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.

**Uncorrelated Coefficients**

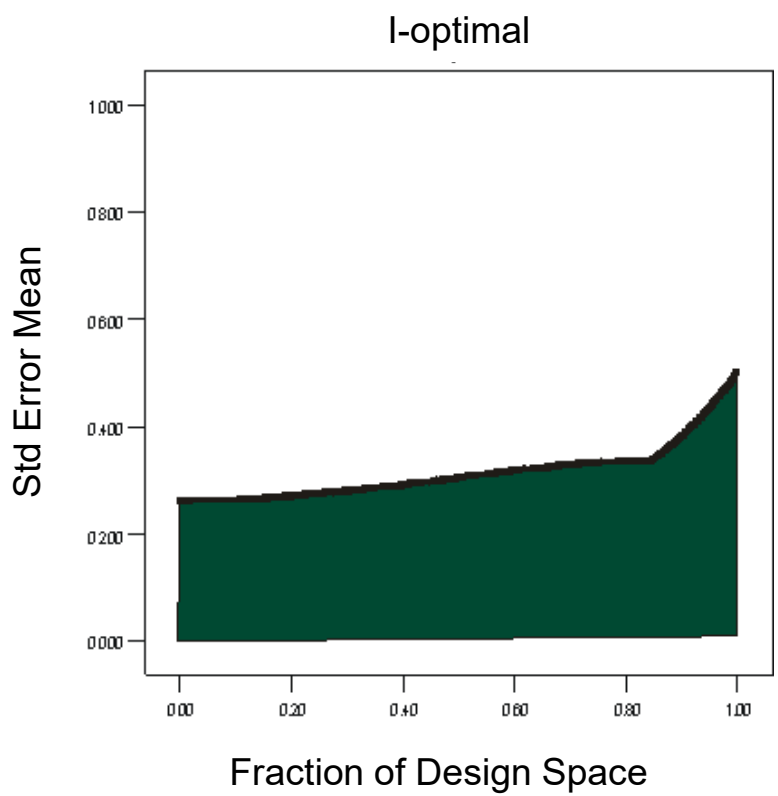



**Correlated Coefficients**



# Optimal Point Selection
## I-optimal Point Selection

An **I-optimal** design seeks to minimizes 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.



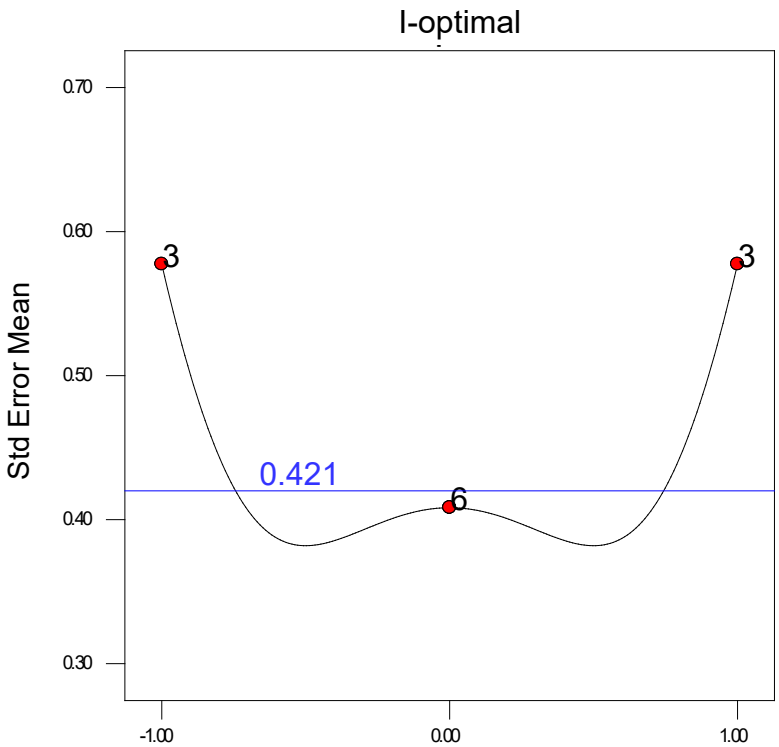
# Optimal Point Selection
## I versus D Optimal Design

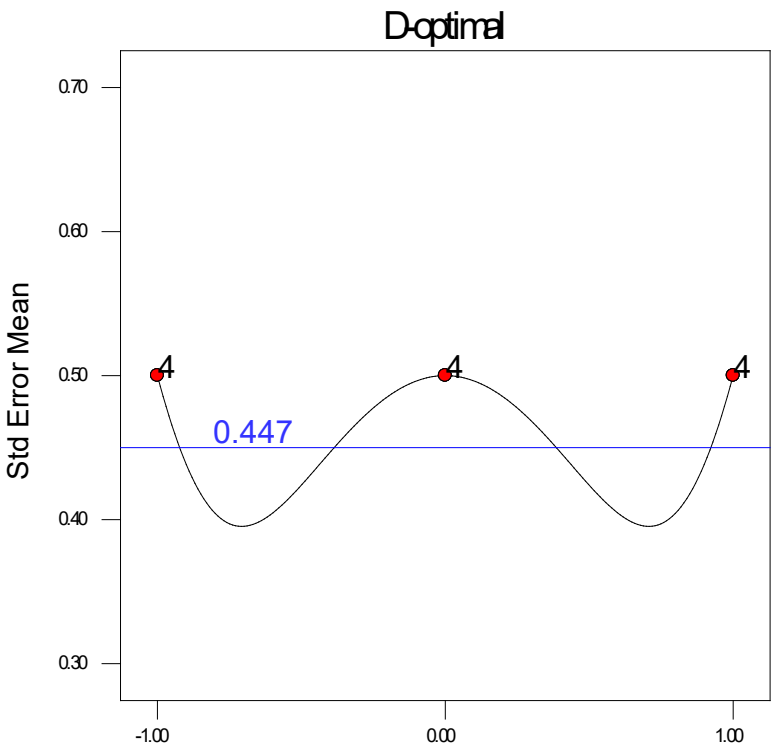
Compare point selection using I-optimal and D-optimal:

- Build a one factor design.
- Design for a quadratic model.
- Choose all twelve runs using optimality as only criterion.

# I versus D Optimal Design

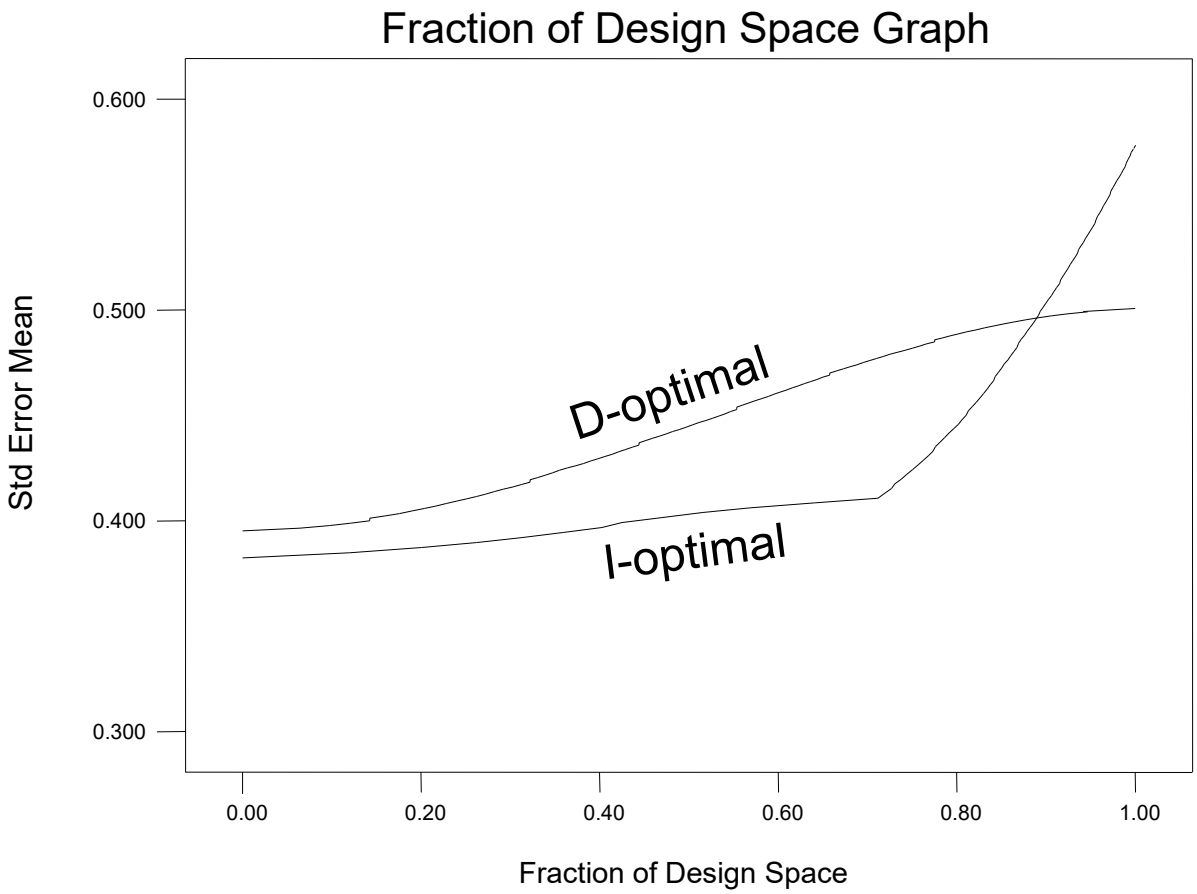
## Optimal 12 Point Designs



# I-optimal versus D-optimal

## One Factor 12 Optimal Points




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

# What about G-Optimality?

## Three 6-point 2-factor Designs

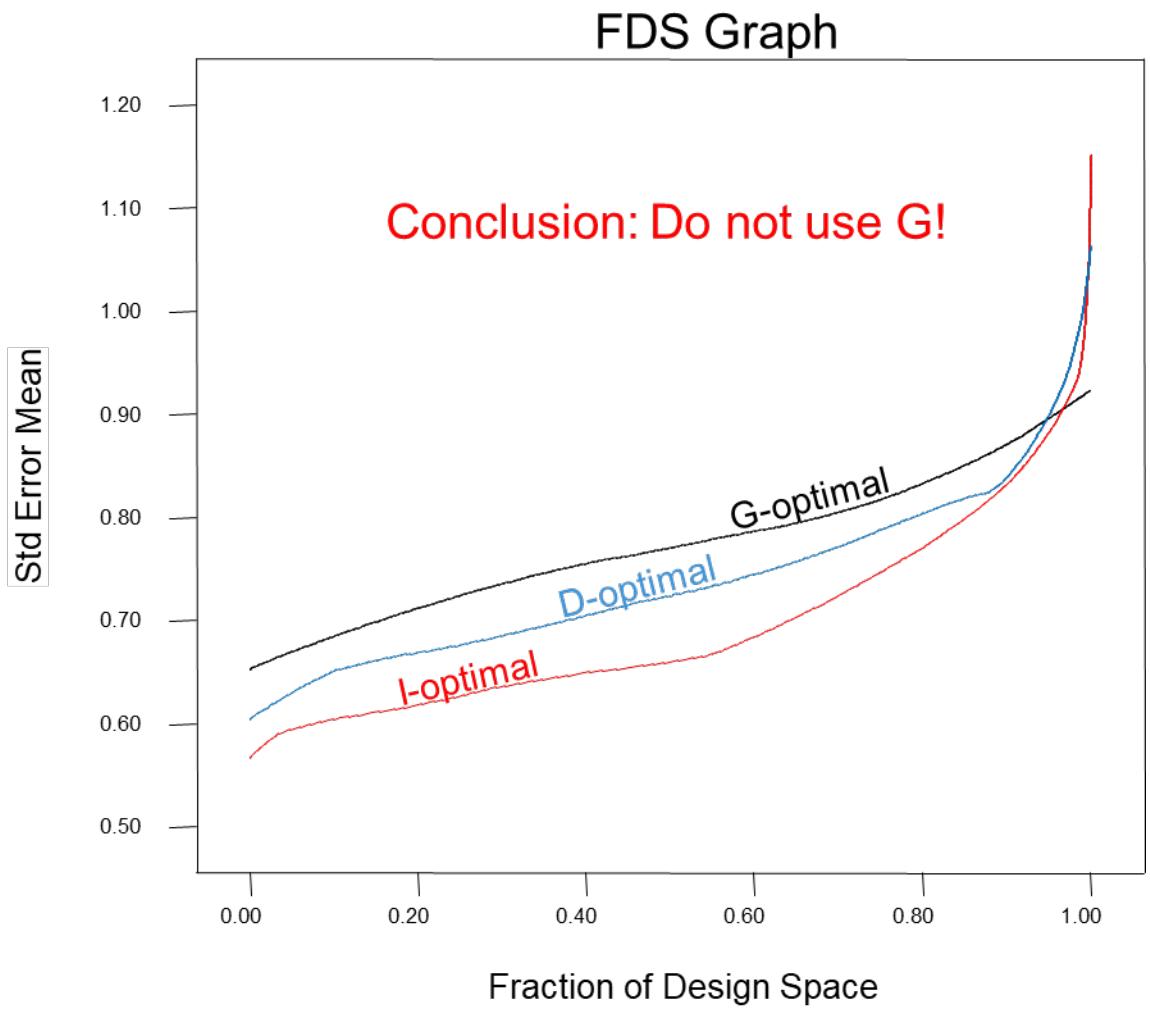
| | G-optimal | D-optimal | I-optimal |
|---|---|---|---|
| **G efficiency** | 87.9% | 66.4% | 56.5% |
| **Min SE mean** | 0.653 | 0.604 | 0.566 |
| **Ave SE mean** | 0.777 | 0.743 | 0.699 |
| **Max SE mean** | 0.923 | 1.063 | 1.152 |

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



# Optimal Point Selection
## I versus D Optimal Design

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!**

# Agenda

Using Optimal Designs to Solve Practical Experimental Problems:

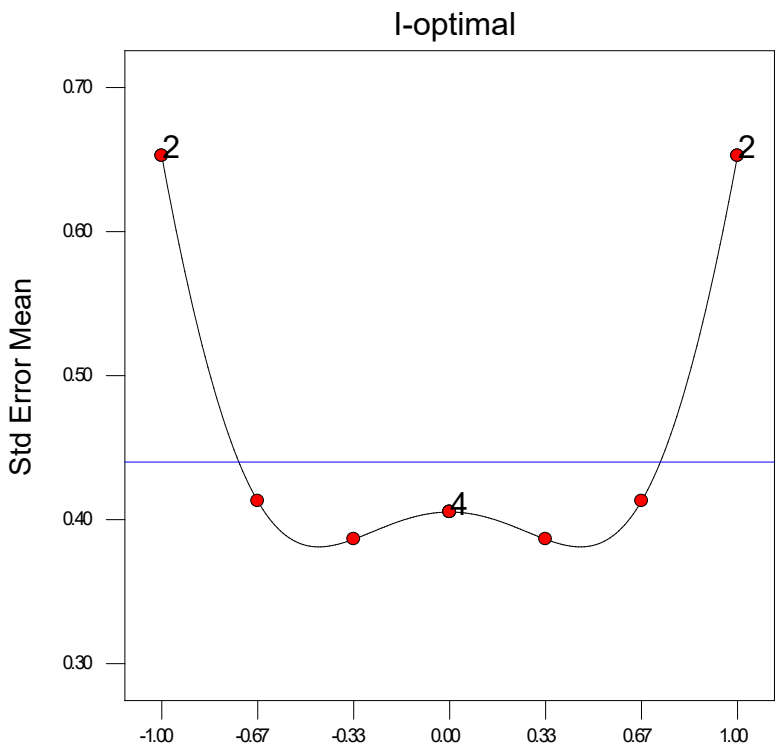
- What's required for a good design.
- Optimal point selection (*I versus D optimality*).
- **Practical aspects algorithmic design.**
- Optimal design example.
- Conclusion and recommendations.

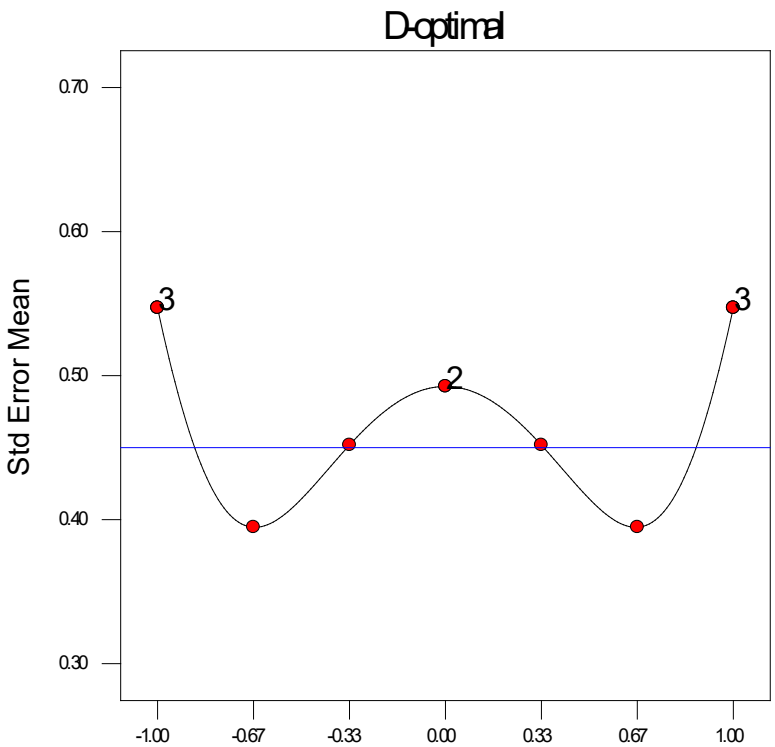
# Optimal Point Selection
## I versus D Optimal Design

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.)*

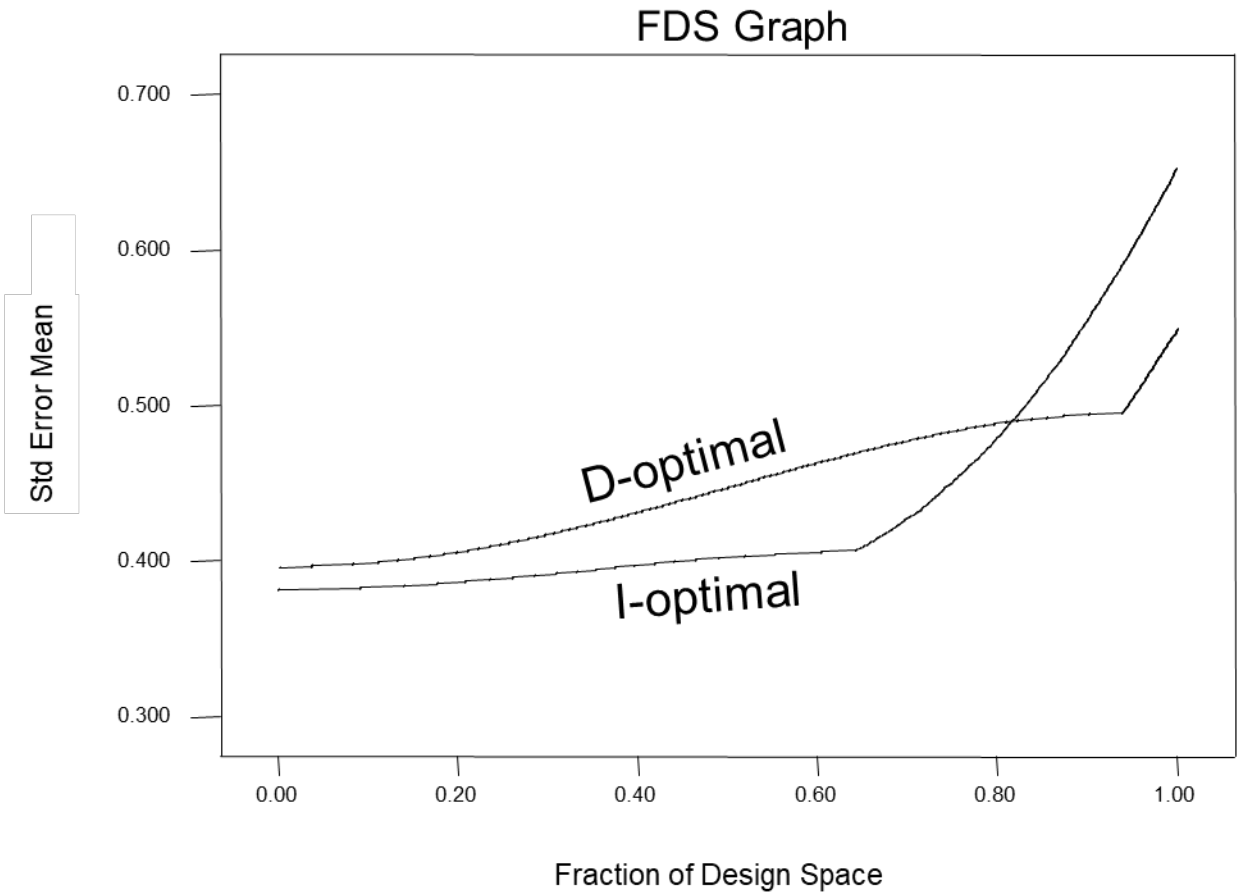
# Optimal Designs
## 8 Optimal + 4 LOF Points



# I-optimal versus D-optimal

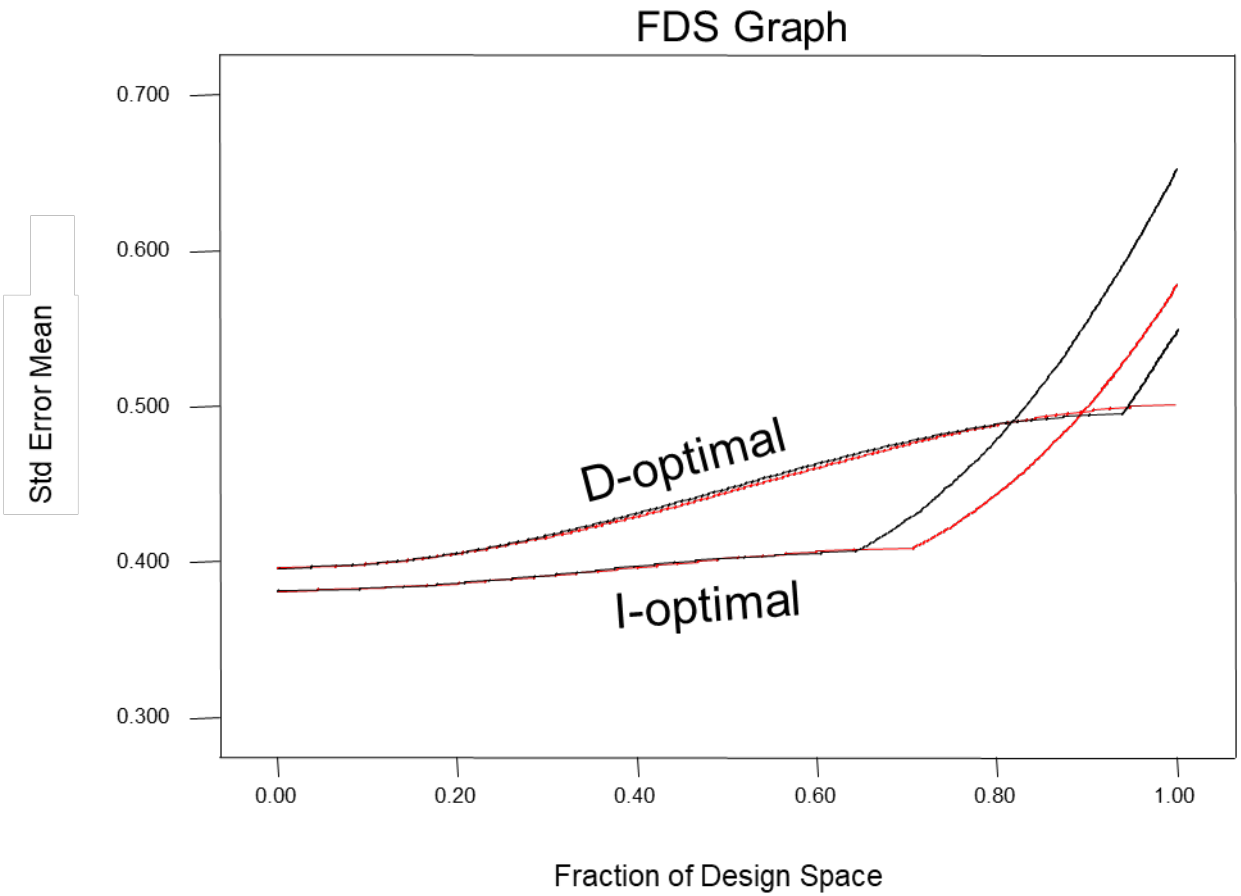
## One Factor 8 Optimal and 4 Distance Points




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

# I-optimal versus D-optimal

## 12 Optimal - 8 Optimal + 4 Distance Points




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 |

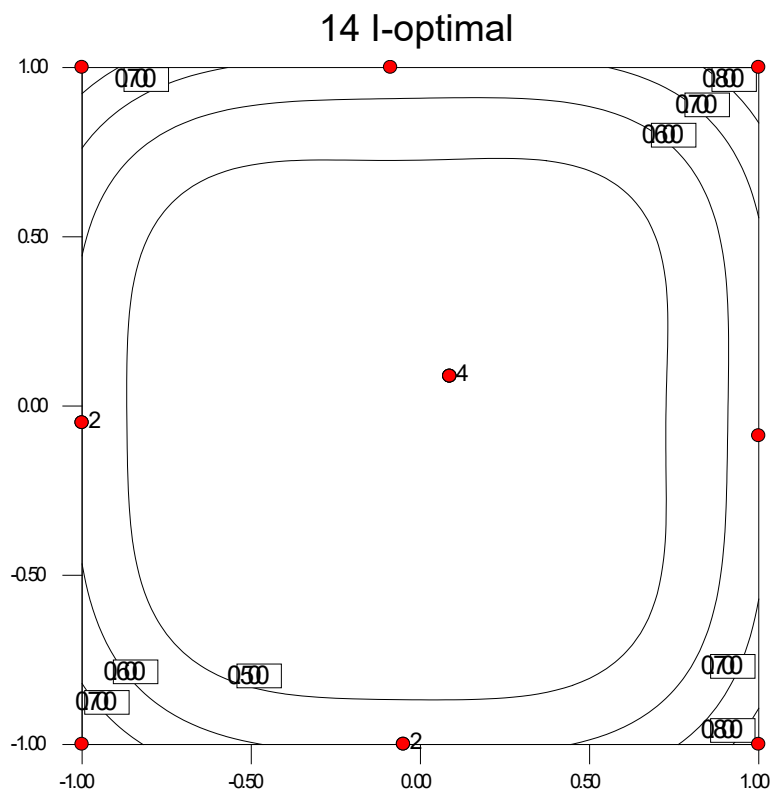
# Optimal Point Selection
## I versus D Optimal Design

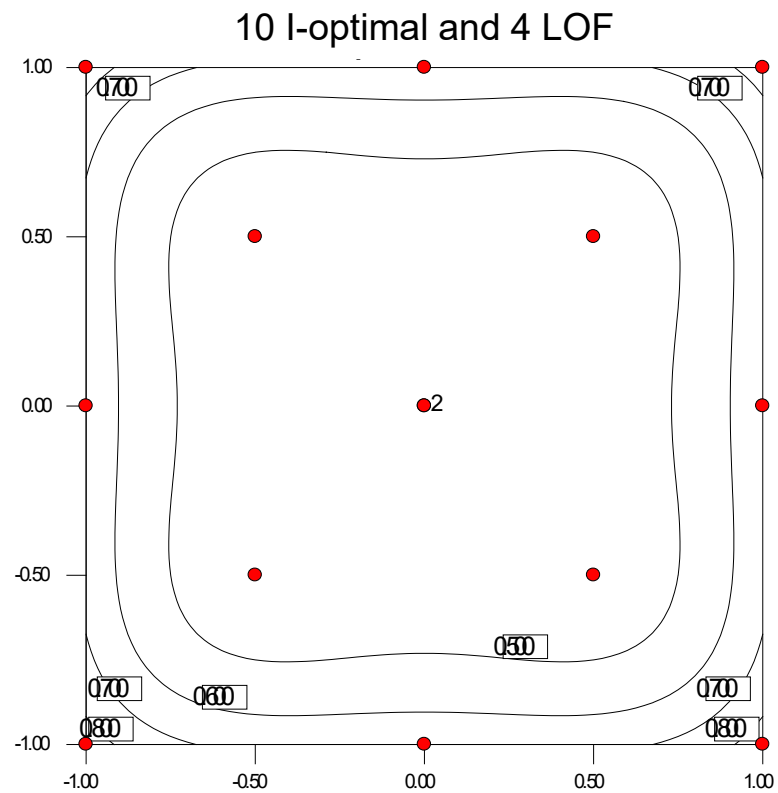
Compare point selection for a two-factor 14-run design:

- 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

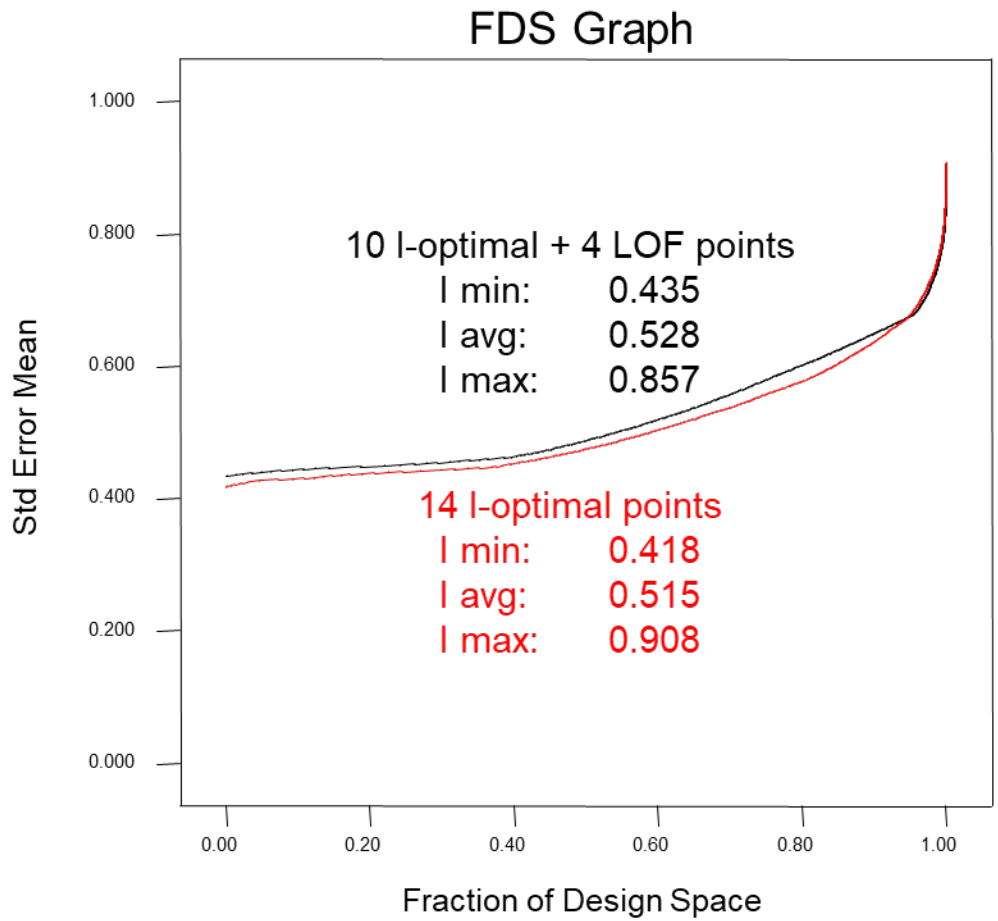
## 14 Run Designs with 0 and 4 LOF Points



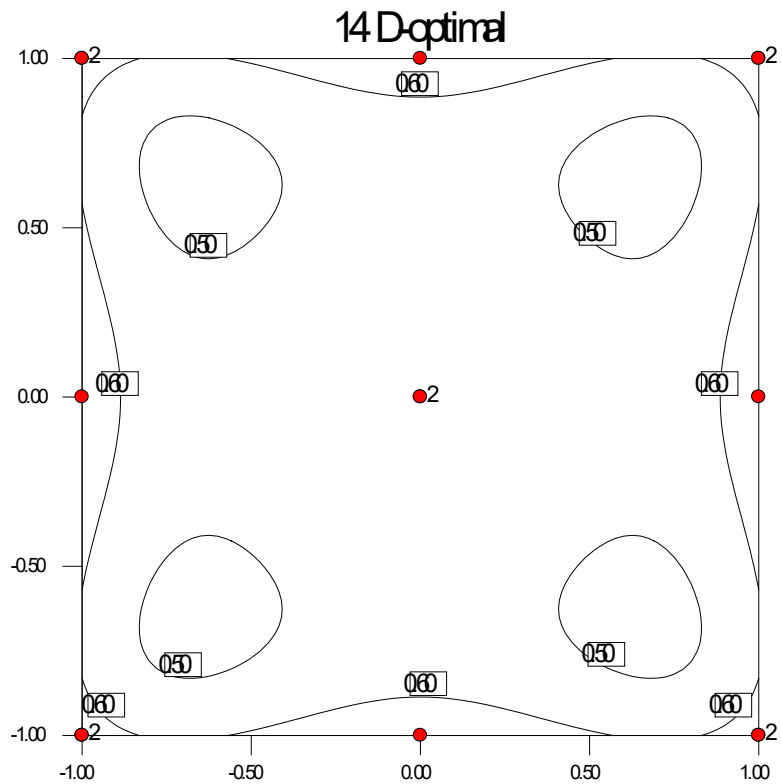
# I-optimal Designs

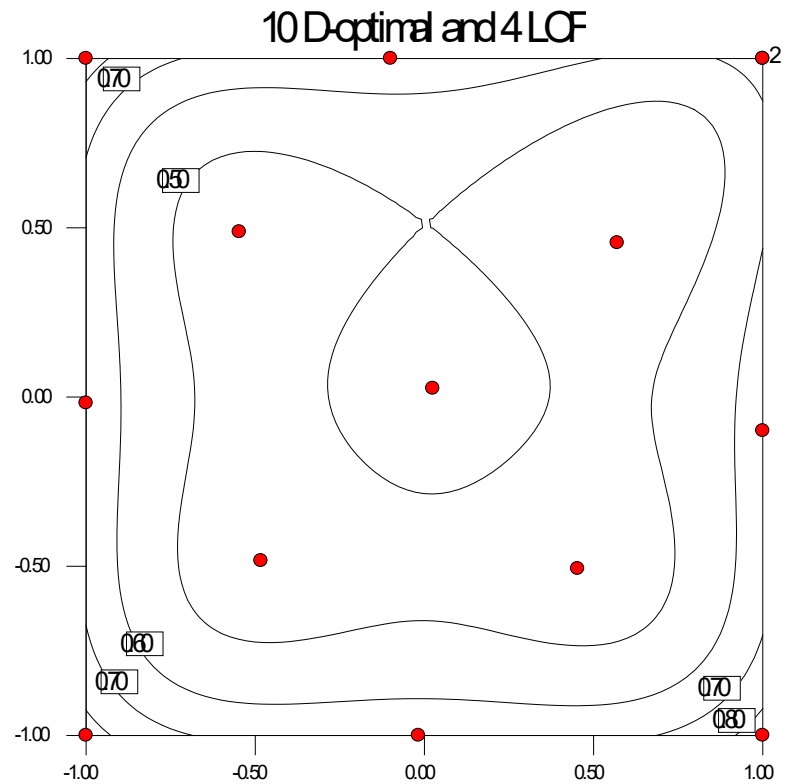
## 14 Run Designs with 0 and 4 LOF Points



# D-optimal Designs

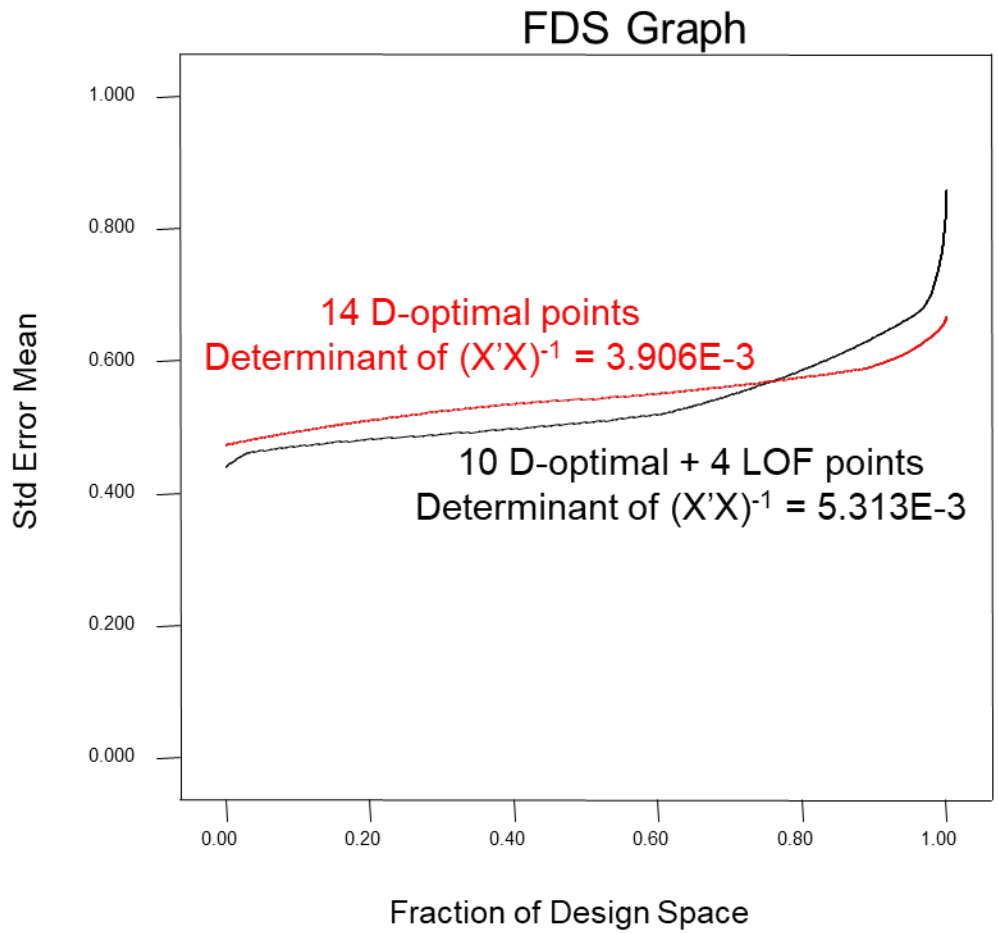
## 14 Run Designs with 0 and 4 LOF Points



# D-optimal Designs

## 14 Run Designs with 0 and 4 LOF Points



# Practical Aspects Algorithmic Design

## Lack of Fit Points

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!

# Practical Aspects Algorithmic Design

## Pure Error Estimation

Estimating pure error:

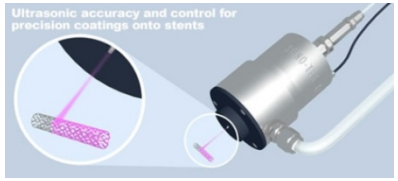
- In physical experiments it is desirable build in an estimate of experimental error.
- Replicates provide an estimate of experimental error independent of model assumptions.
- Adding replicates is a good trade off!

# 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)




| | Name | Units | –1 level | +1 level |
|---|---|---|---|---|
| A | flow rate | ml/min | 10 | 30 |
| B | pressure | kPa | 3 | 10 |
| C | linear speed | inch/sec | 0.1 | 0.5 |

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.

# Prevent "not enough" Coating

(A = 10, B = 3 and C = 0.5)

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) ≥ 15 when B = 3 and C = 0.5

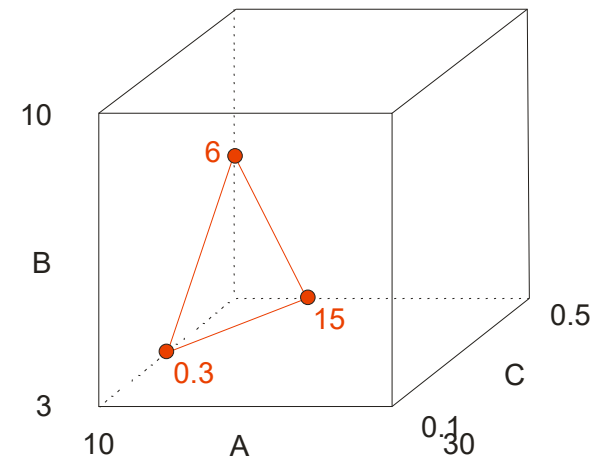
  **$CP_A = 15$**

- B (pressure) ≥ 6 when A = 10 and C = 0.5

  **$CP_B = 6$**

- C (linear speed) ≤ 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) ≤ 20 when B = 10 and C = 0.1

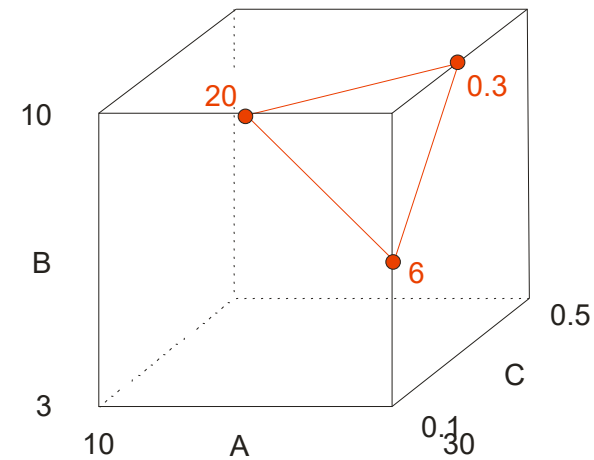
  **$CP_A = 20$**

- B (pressure) ≤ 6 when A = 30 and C = 0.1

  **$CP_B = 6$**

- C (linear speed) ≥ 0.3 when A = 30 and B = 10

  **$CP_C = 0.3$**

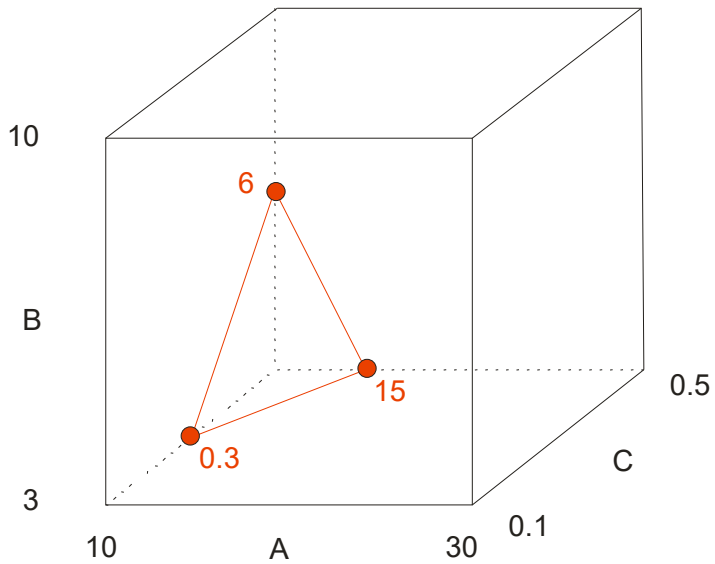


# Prevent Not enough and Too much

## Multiple Linear Constraints

### Not Enough

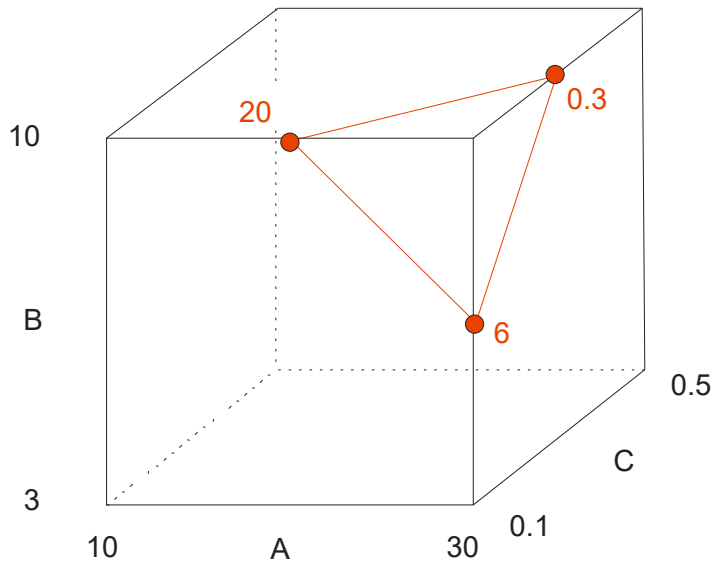
exclude (10, 3, 0.5)




$4.5 \leq 0.6A+B-15C$

### Too Much

exclude (30, 10, 0.1)




$0.8A+2B-40C \leq 32$

# Practical Aspects Algorithmic Design

## One Suggestion for Point Selection

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:

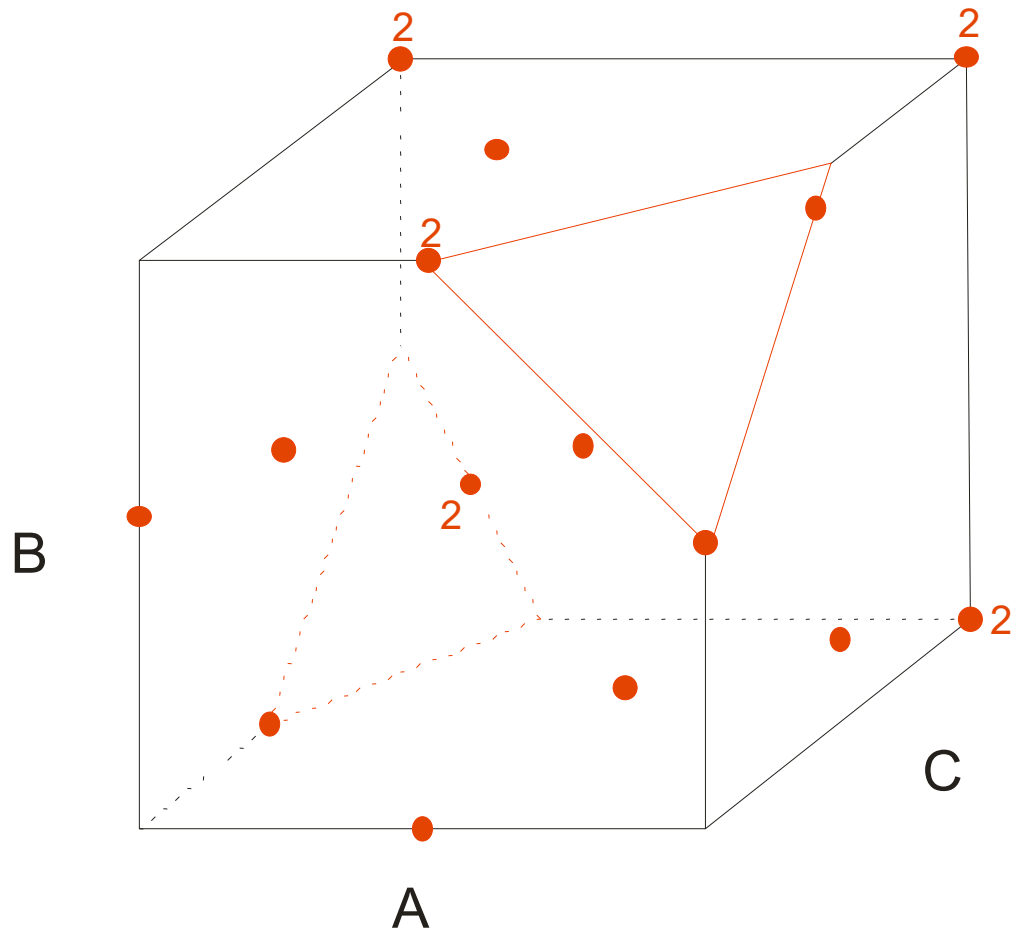
- Model: p points using an optimality criterion
- Lack-of-Fit: 5 points; based on distance or estimating higher order model terms.
- Replicates: 5 points, using the model optimality criterion (most influential).

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



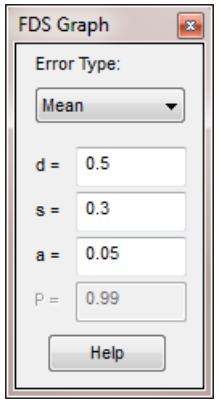
# Spray Coating
## Evaluate your I-optimal Design

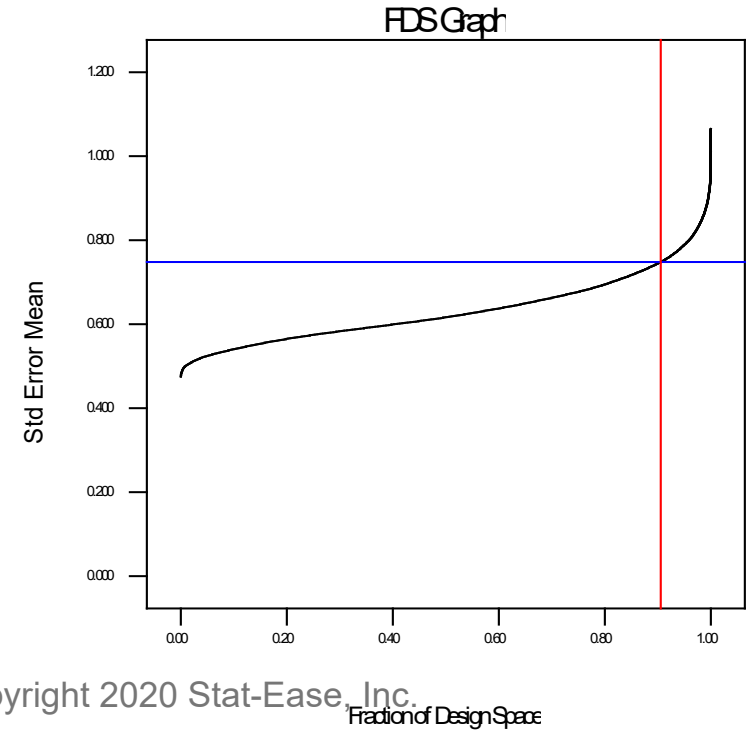
Is the optimal design precise enough?

- Want the estimated mean of the “average thickness” (thickness) to be within ± 0.5.
- The estimated standard deviation is 0.30.

Design-Expert® Software

Min Std Error Mean: 0.475
Avg Std Error Mean: 0.637
Max Std Error Mean: 1.065
Constrained
Points = 50000
t(0.05/2,10) = 2.22814
FDS = 0.91
Std Error Mean = 0.748



# 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.**

# RSM Design Summary

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!

# Practical Aspects of DOE

## Remember what is Most Important

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).

# Practical Aspects Algorithmic Design
## Optimality Criteria

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

# Practical Aspects Algorithmic Design

## Suggestion for Point Selection

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


[1] Christine M. Anderson-Cook (2010), "A Matter of Trust", *Quality Progress*, March 2010.

[2] Raymond H. Myers and Douglas C. Montgomery (2002), 2nd edition, *Response Surface Methodology*, John Wiley and Sons, Inc.

[3] George E. P. Box and Norman R. Draper (2007), *Response Surfaces, Mixtures and Ridge Analyses*, John Wiley and Sons, Inc.

[4] Alyaa R. Zahran, Christine M. Anderson-Cook and Raymond H. Myers, "Fraction of Design Space to Assess Prediction", *Journal of Quality Technology*, Vol. 35, No. 4, October 2003.

[5] Heidi B. Goldfarb, Christine M. Anderson-Cook, Connie M. Borror and Douglas C. Montgomery, "Fraction of Design Space plots for Assessing Mixture and Mixture-Process Designs", *Journal of Quality Technology*, Vol. 36, No. 2, October 2004.

[6] Myrta Rodriguez, Bradley Jones, Connie M. Borror, and Douglas C. Montgomery, "Generating and Assessing Exact G-Optimal Designs", *Journal of Quality Technology*, Vol. 42, No. 1, January 2010.

[7] Patrick Whitcomb and Gary Oehlert, "Sizing Mixture (RSM) Designs", Chemical and Process Industries web site: http://www.asq.org/divisions-forums/cpi/.

[8] Mark J. Anderson and Patrick J. Whitcomb, "Practical Aspects for Designing Statistically Optimal Experiments", *Journal of Statistical Science and Application*, 2 (2014) 85-92, www.statease.com/pubs/practical-aspects-for-designing-statistically-optimal-experiments.pdf.

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

**Modern DOE for Process Optimization**

**Mixture Design for Optimal Formulations**

| Individuals | Teams (6+ people) |
|---|---|
| Improve your DOE skills | Choose your date & time |
| Topics applicable to both novice and advanced practitioners | Add company case studies |

Learn more: **www.statease.com**

Contact: **workshops@statease.com**

StatEase®

statistics made easy®

*Make the most from every experiment!℠*