How many runs do I need? Using power and precision to size DOE's

Shari Kraber, MS Applied Statistics
Senior Client Success Manager

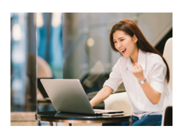
shari@statease.com

Licensing questions? sales@statease.com

December 2021

Maximizing this educational opportunity

Welcome everyone! To make the most from this webinar:



- Attendees on mute
- Questions addressed afterward
- Send further questions to shari@statease.com

PS: Presentation posted to www.statease.com/webinars/

Please press the raise-hand button if you are with me.

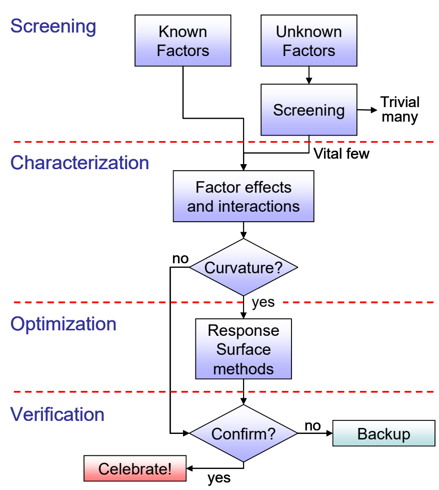
Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Agenda: Using Power & Precision to Size DOEs

- **Overview of sizing designs based on objectives**
- Using power to size factorial designs
  - Factorial case study
- Using precision to size response surface designs
  - RSM case study
  - Mixture case study
- Summary

Strategy of Experimentation*

*webinar: *Know the SCOR for Multifactor Strategy of Experimentation*



Design Objective

Detect main effects

↓

Define interactions

↓

Make precise predictions

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Differences for Sizing Designs
Factorial versus RSM/Mix

Factorials	Response Surface/Mixtures
Focus: screening and characterization to identify main factor effects and interactions; respectively.	Focus: modeling a response surface to optimize and make predictions.
What are the important process factors?	How well does the surface represent true behavior?
For this purpose, **power** is an ideal metric to evaluate design suitability, and determine an appropriate number of runs.	For this purpose, **precision** is a better measure to ensure the experiment design is sized correctly.

Agenda: Using Power & Precision to Size DOEs

- Overview of sizing designs based on objectives
- **Using power to size factorial designs**
  - Factorial case study
- Using precision to size response surface designs
  - RSM case study
  - Mixture case study
- Summary

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

DOE Process *(1 of 2)*

1. Identify opportunity and define objective.
2. State objective in terms of measurable responses.
   a. Define the minimal change (Δy*) that is <u>important</u> to detect for each response (signal).
   b. Estimate experimental error (σ) for each response (noise).
   c. Use the signal to noise ratio (Δy/σ) to estimate power.

**See next slide for tips on defining your signal.*

DOE Process
Defining Your Signal (Δy)

To define your signal, ask this question:

"What is the <u>minimum</u> amount of change in the response that will be recognized as an important improvement?"

The answer is a <u>business decision</u>, not a statistical calculation.

Objective: Improve yield from the current level of 80%. Each percent is worth $100,000 per year in profits.

Signal: What amount of improvement will be valued?

0.1% | 1% | 10%

A <u>quantitative</u> answer is required!

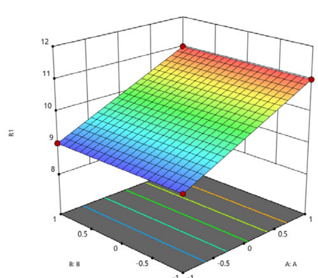
Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

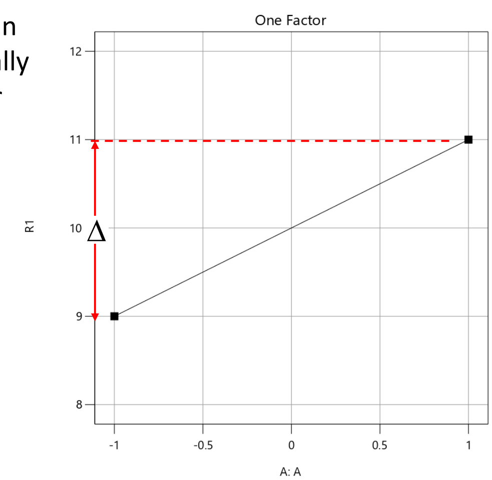
DOE Process *(2 of 2)*

3. Select the input factors and ranges to study. (*Choose factor ranges that are likely to change the response by at least Δy.*)
4. Select a design and:
   - Evaluate aliases.
   - Assess power.
   - Examine the design layout to ensure all the factor combinations are safe to run and likely to result in meaningful information (no disasters).

Factorial Design – Power
2^3 Full Factorial Δ=2 and σ=1

Define the smallest change in the response that is practically important to you (or to your customer).





Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Factorial Design – Power
2^3 Full Factorial Δ=2 and σ=1

1 replicate (8 runs)

Name	Units	Delta (Signal)	Sigma (Noise)	Signal/Noise	Power for A	Power for B	Power for C
R1		2	1	2	**57.2%**	**57.2%**	**57.2%**

2 replicates (16 runs)

Name	Units	Delta (Signal)	Sigma (Noise)	Signal/Noise	Power for A	Power for B	Power for C
R1		2	1	2	**95.6%**	**95.6%**	**95.6%**

Power is reported at a 5.0% alpha level to detect the specified signal/noise ratio.

Power should be approximately 80% or greater for the effects you want to detect.

What is Power?
No Factor Effect; H_0: Δ = 0

Power = (1−β)*100%

Power is the probability of revealing an active effect of size delta (Δ) relative to the noise (σ) as measured by signal to noise ratio (Δ/σ).

It should be high (at least 80%!) for the effect size of interest.

Effect?		ANOVA says: Retain H_0	ANOVA says: Reject H_0
Truth:	***No***	OK☺	Type I Error (alpha) *False Alarm*
Truth:	***Yes***	Type II Error (beta) *Failure to detect*	OK☺

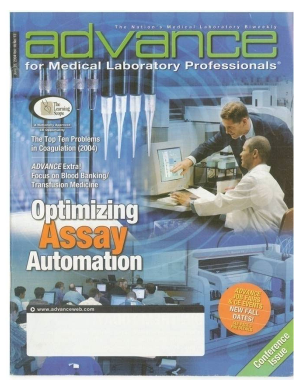
Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Agenda: Using Power & Precision to Size DOEs

- Overview of sizing designs based on objectives
- **Using power to size factorial designs**
  - **Factorial case study**
- Using precision to size response surface designs
  - RSM case study
  - Mixture
- Summary

Application of DOE to Mouse Cell Assay*

This case study highlights how a development team applies DOE to study a mouse-cell fluorescent assay performed in a 96-well plate format. They are concerned about the effects of several key factors.



** Detailed in "How Experimental Design Optimizes Assay Automation" by Thomas Erbach & Lisa Fan, Beckman Coulter, Inc., Shari Kraber, Stat-Ease, Inc.,* Advance, *June 28, 2004, Vol. 16, No. 13,pp 18-21.*

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Mouse Cell Assay

DOE Process *(page 1 of 3)*

1. Identify opportunity and define objective.
   The objective is to maximize signal from the assay.
2. State objective in terms of measurable responses.
   a. Define the change (Δy) that is important to detect.
      A difference of 400 fluorescent units is of interest;
      $\Delta y \approx 400$.
   b. Estimate experimental error (σ) for each response.
      Historical data is used to estimate the standard deviation;
      $\sigma \approx 400$.
   c. Use the signal-to-noise ratio ($\Delta y/\sigma$) to estimate power.
      $\Delta y/\sigma = 400/400 =$ **1.0**

Mouse Cell Assay

DOE Process *(page 2 of 3)*

3. Select the input factors and ranges to study. (*Choose factor ranges that are likely to change the response by at least Δy.*)

Factor	−1 level	+1 level
A. Cell Number	5000	10000
B. Stimulant	5 μL	10 μL
C. Substrate concentration	0.15 μM	0.30 μM

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Mouse Cell Assay
DOE Process *(page 3 of 3)*

4. Select a design (a full 2^3 two-level factorial) and evaluate:
   - Aliases (fractional factorials and/or blocked designs) Not an issue with this design choice (running all combinations).
   - All factor combinations for safety and reasonability (likelihood of producing meaningful information).
     Assume the team knows from subject matter expertise and actual range-finding tests that all runs will be do-able and informative.
   - Power (ideally at least 80% probability for detection).
     See following statistics on main-effect estimates (anticipating sparsity of effects)

Mouse Cell Assay
Replicated Factorial Design

For this study, the experimenter wants to replicate this 2^3 design, but how many replicates are enough?

Evaluating "Power" gives us the answer!

Mouse.dxpx
Rebuild,
showing power

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

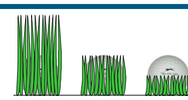
Mouse Cell Assay
Evaluating Power

Power to detect main effects at 5% alpha:

- 8 unique design runs: 19.5% ☹
- 16 runs – 2^{nd} replicate of original 2^3: 45.2% ☹
- 24 runs – 3^{rd} replicate: 64.5% 😐
- 32 runs – 4^{th} replicate : 77.9% 😐
- 40 runs – 5^{th} replicate : 86.8% ☺

Less than half of the 96 wells suffice for adequate power (80%). Further runs provide greatly diminishing returns.

Power
The probability of finding an effect!



Power depends on:

- The size of the difference Δ:
  the larger the difference the higher the power.
- The size of the experimental error σ:
  the smaller σ the higher the power.
- Choice of design being appropriate to the problem:
  larger designs have more power.
- The number of replicates:
  the more runs the higher the power.

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Agenda: Using Power & Precision to Size DOEs

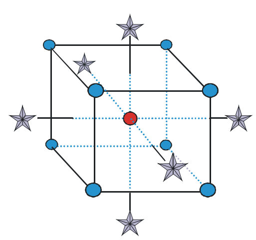
- Overview of sizing designs based on objectives
- Using power to size factorial designs
  - Factorial case study
- **Using precision to size response surface designs**
  - RSM case study
  - Mixture case study
- Summary

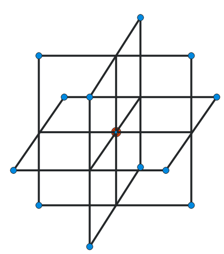
Response Surface Designs*

*webinar: *Keys to Building Response Surface Designs*

All these designs are statistically sound, and likely more efficient than what an experimenter would create by hand.

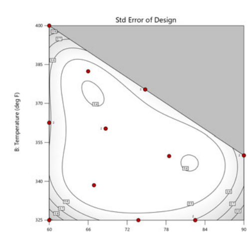
Central Composite





Box-Behnken

Optimal



Goal: Model the design space well to obtain precise predictions

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

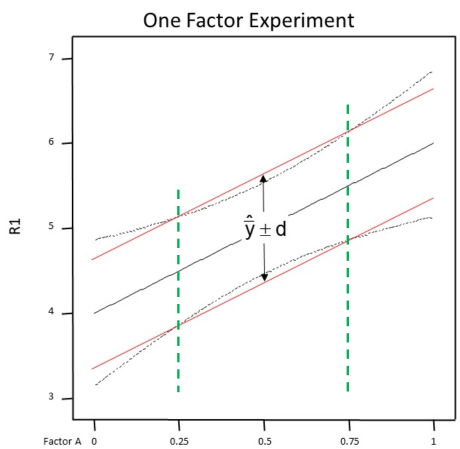
RSM Goal: Making Precise Predictions

The **precision** of prediction depends on the:

- **location** in the design space
- **standard deviation** "s" of the response.

Notice how the confidence bands (black dotted) vary over this one-factor response surface.

In this case, only about 50% of the **fraction of design space** falls within the desired halfwidth "d" (red).



To right size an RSM, do enough runs to achieve ≥80% FDS.

The Inputs for Sizing via FDS

Precision (d): How well do you want to make predictions? The more precision you want, the more data is required. This is a business decision.

"We want to estimate the mean response with a precision (d) of +/- 0.80."

Standard Deviation (s): This is the process standard deviation (including sampling and test variation). It is typically estimated from historical data, prior DOEs or other means. The greater the standard deviation, the more data is required. (Same as power.)

Historical data provides a standard deviation of 0.50.

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Examples for “d” and “s”

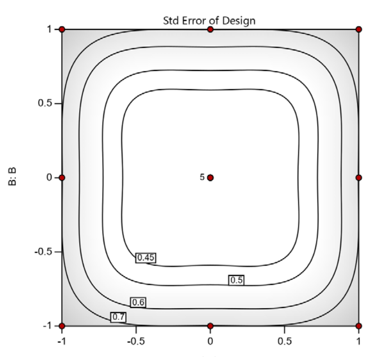
Response	Desired Precision (d)*	Standard Deviation (s)**
Viscosity	$\hat{Y}$ +/- 0.15 cp	0.12 cp
Chemical conversion	$\hat{Y}$ +/- 5%	4%
Flex modulus	$\hat{Y}$ +/- 4 psi	3.7 psi
Avg thickness	$\hat{Y}$ +/- 4.5 mm	3 mm

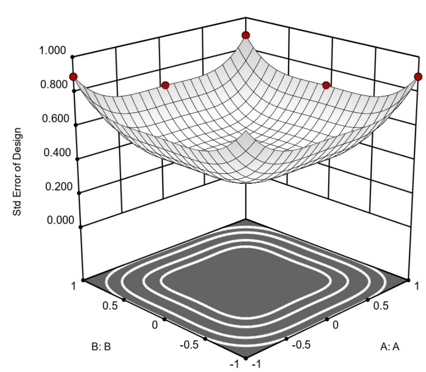
* Precision – a business decision
** Standard deviation – generally calculated from historical data

Tool to Assess Precision

Standard Error Plot: Two-factor FCD *(1 of 2)*

This plot shows the standard error of the predictions for a face-centered central composite design (FCD). Error is lower in the middle and higher near the edges and corners. *(Evaluation – Graphs – Contour)*

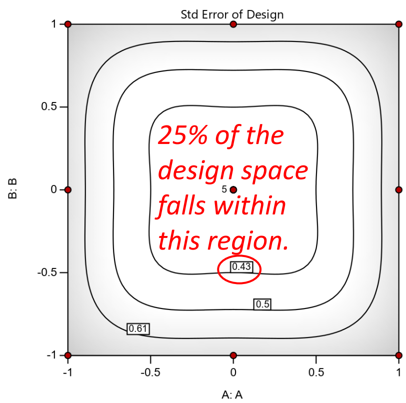


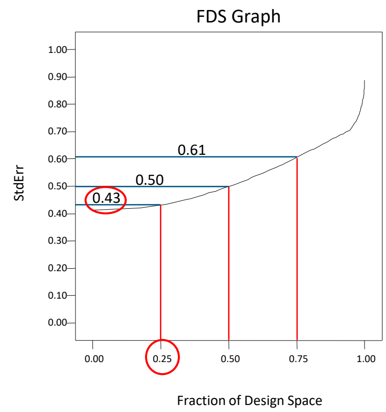


Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Standard Error (SE) Plot → FDS *(2 of 2)*

The fraction of design space (FDS) graph provides a **profile** of the prediction error across the design space. In this case 25% of the design space—the inner core—estimates SE ≤0.43 and so on.





Agenda: Using Power & Precision to Size DOEs

- Overview of sizing designs based on objectives
- Using power to size factorial designs
  - Factorial case study
- **Using precision to size response surface designs**
  - **RSM case study**
  - Mixture case study
- Summary

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Response Surface Method Case

Help, Tutorials: Response Surface

This case study on a chemical process features two key responses:

1. Conversion (%)
2. Activity

The engineers tested three process factors:

A. Time (minutes)
B. Temperature (degrees C)
C. Catalyst (percent)

They ran a CCD 20 runs in two blocks, machine-by-machine, divided into:

- 8 factorial points with 4 center points (12 runs in total), and
- 6 axial (star) points with 2 more center points (8 runs).

Response Surface Methods Case

RSM DOE Process *(page 1 of 2)*

1. Identify opportunity and define objective.
   - Maximize conversion to be >80
   - Find conditions that target Activity at 63
2. State objective in terms of measurable responses.
   - Define the precision (d) needed.
     - Conversion ± 5.0%
     - Activity ± 1.3
   - Estimate standard deviation (σ) for each response.
     - $s_{conversion} \approx 4.0$
     - $s_{activity} \approx 1.1$
   - Calculate the d/s ratios:
     - Conversion: 5/4 = 1.25
     - Activity: 1.3/1.1 = 1.18 ← worst-case scenario

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Response Surface Methods Case

RSM DOE Process *(page 2 of 2)*

3. Select the input factors and ranges to study. (*Consider both your region of interest and region of operability.*)

   40 to 50 minutes, 80° to 90°C, and 2 to 3% catalyst

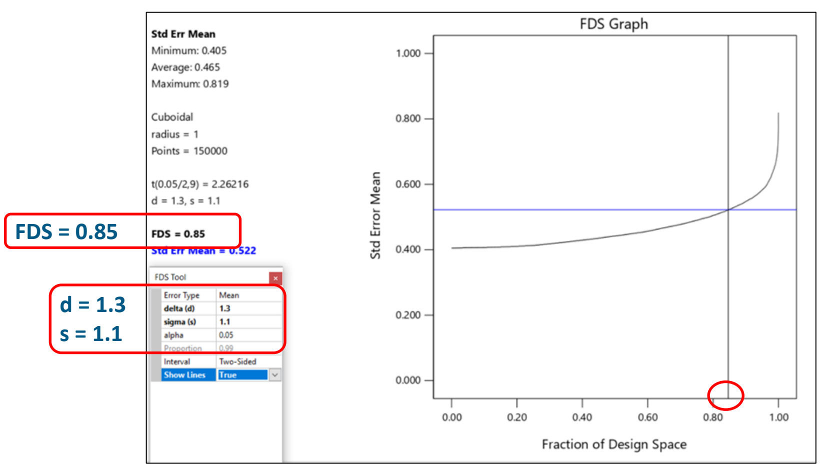
4. Choose the polynomial to estimate. Quadratic

5. Select a design (Central Composite) and:
   - Size design for precision needed.
   - Examine the design layout to ensure all the factor combinations are safe to run and are likely to result in meaningful information (no disasters).

RSM.dxpx
Rebuild,
showing precision

Response Surface Methods Case

Evaluation – Graphs – FDS



For this worst-case, 85% of the design space will make predictions with desired precision. ☺ (Conversion will generate even better FDS.)

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Sizing for Precision
What Level of FDS is Good Enough?

Rule of thumb for precise predictions: **FDS ≥ 80%**
(Easy to remember--also the goal for power.)

What can be done to improve the FDS?

- Manage expectations; i.e., increase d
- Decrease noise; i.e., decrease s
- Increase the number of runs in the design. For optimal designs this can be easily done by rebuilding with additional model points.

Agenda: Using Power & Precision to Size DOEs

- Overview of sizing designs based on objectives
- Using power to size factorial designs
  - Factorial case study
- **Using precision to size response surface designs**
  - RSM case study
  - **Mixture case study**
- Summary

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Optimal Flare

4 Component I-Optimal Mix DOE

In manufacturing a particular type of flare, the chemical constituents are: A - magnesium, B - sodium nitrate, C - strontium nitrate, and D - binder. Experience dictates the following constraints:

$$0.40 \le A \le 0.60$$
$$0.10 \le B \le 0.50$$
$$0.10 \le C \le 0.50$$
$$0.03 \le D \le 0.08$$
$$\text{Total} = 1.00$$

- The problem is to find the blend (A, B, C, D) which gives the maximum illumination (light), measured in candles.
- Experience suggests a special cubic model for illumination.

Optimal Flare

1. Chose an Optimal mixture and enter the four components and their constraints and a total of 1:

Optimal (Custom) Design

A flexible design structure to accommodate custom models, categoric factors, and irregular (constrained) regions. Runs are determined by a selection criterion chosen during the build.

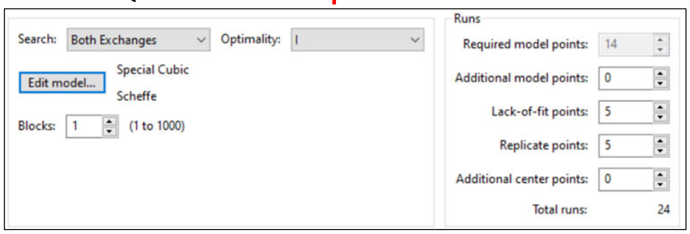
Mixture components: 4 (2 to 24) Total: 1 Units: ◉ Horizontal ○ Vertical

	Name	Low	High
A [Mixture]	Magnesium	0.4	0.6
B [Mixture]	Na nitrate	0.1	0.5
C [Mixture]	Sr nitrate	0.1	0.5
D [Mixture]	binder	0.03	0.08

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Optimal Flare

2. Use "**Both Exchanges**" to build an "**I-optimal**" design and change the default "Quadratic" to a "**Special Cubic**" model:



3. One response "**illumination**" with units of "**candles**".

***Hint**: Be sure to change Quadratic to Special Cubic model.*

Optimal Flare
Sizing for precision with FDS

Does the design with 24 runs have adequate precision?

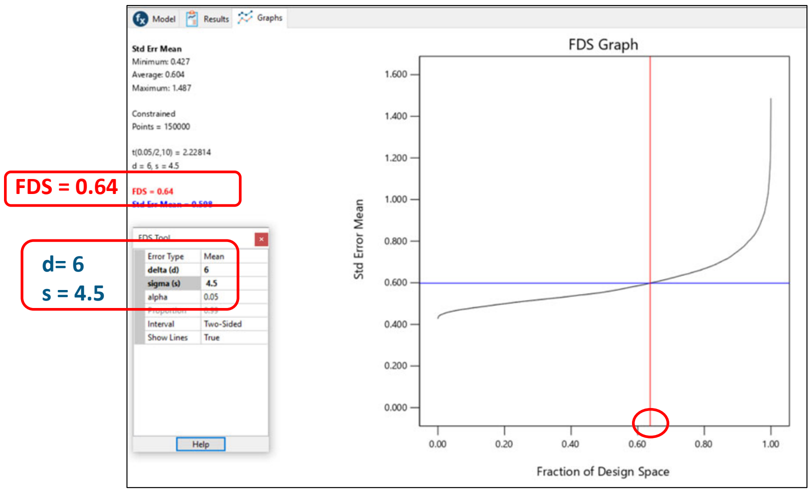
- Determine d:
  - Want FDS ≥ 80% with precision of **± 6**
- Determine s:
  - Std. dev. for illumination (estimated from SPC data) is **4.5**

Flare design 1.dxpx
Rebuild,
showing precision

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

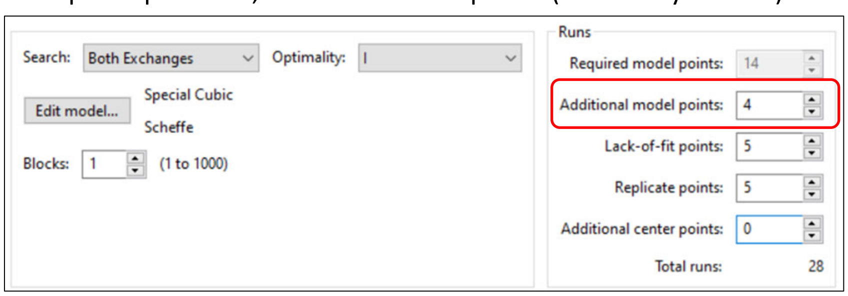
Optimal Flare
Evaluation – Graphs – FDS

Only 64% of the design space has StdErr ≤ 0.598



Optimal Flare
Sizing for precision with FDS

To improve precision, add more model points (we will try 4 more):



Increasing model points is the most effective way to improve precision. Assuming there are already 5 each for lack of fit and replicates, then increase the number of model points.

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

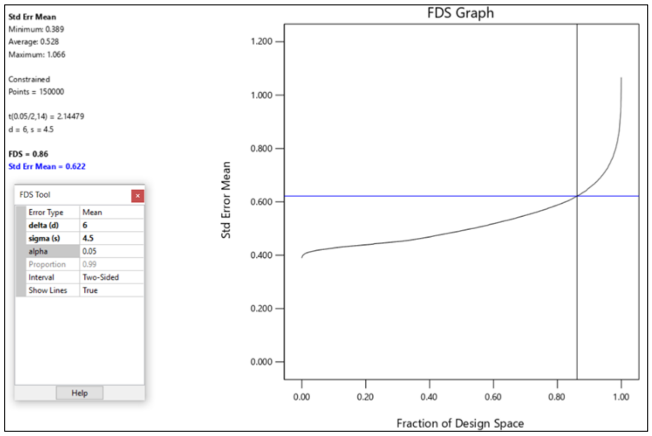
Optimal Flare

Sizing for precision with FDS

Does the design with 28 runs have adequate precision?

FDS = 0.86

86% of the design space will predict to desired precision.



Agenda: Using Power & Precision to Size DOEs

- Overview of sizing designs based on objectives
- Using power to size factorial designs
  - Factorial case study
- Using precision to size response surface designs
  - RSM case study
  - Mixture case study
- **Summary**

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Differences for Sizing Designs
Factorial versus RSM/Mix

Factorials	Response Surface/Mixtures
Focus: screening and characterization to identify main factor effects and interactions; respectively.	Focus: modeling a response surface to optimize and make predictions.
What are the important process factors?	How well does the surface represent true behavior?
For this purpose, **power** is an ideal metric to evaluate design suitability, and determine an appropriate number of runs.	For this purpose, **precision** is a better measure to ensure the experiment design is sized correctly.

Using Power & Precision to Size DOEs
Summary

1. Size your DOE appropriately for its type and purpose:
   - Factorial Designs – size via Power.
     The power to detect each individual effect/coefficient is key.
   - RSM/MIX Functional design – size for Precision.
     Focus on the ability to predict the mean response to a defined amount of precision.
2. For Power (factorial designs), define:
   - Δy : minimum difference you want the DOE to detect
   - s: standard deviation excluding factor effects (historical data)
   - Size to achieve power 80% or greater

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

How to Size DOEs with FDS
Summary

3. For Fraction of Design Space (FDS) (RSM and Mixture), define:
   - d: precision of the predicted mean
   - s: standard deviation excluding factor effects (historical data)
   - Size to achieve FDS 80% or greater

Now ***you*** *can answer the question...How many runs do we really need?*

References
Power and FDS

1. 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.
2. 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.
3. Gary Oehlert and Patrick Whitcomb (2001), Sizing Fixed Effects for Computing Power in Experimental Designs, *Quality and Reliability Engineering International*, July 27, 2001.

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Stat-Ease Training:
Sharpen Up Your DOE Skills

Modern DOE for Process Optimization

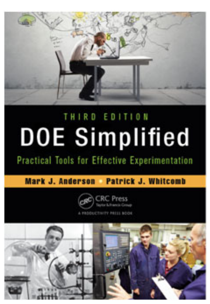
Mixture Design for Optimal Formulations

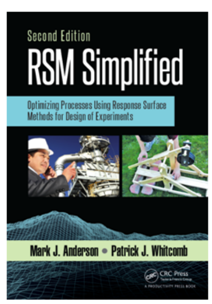
Individuals	Teams (6+ people)
Improve your DOE skills	Choose your date & time
Topics applicable to both novice and advanced practitioners	Add company case studies

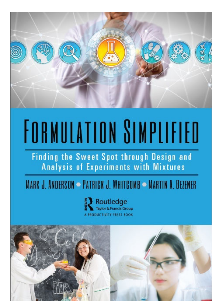
Learn more: **www.statease.com**

Contact: **workshops@statease.com**

Resources*







** Taylor & Francis/CRC/Productivity Press, New York, NY.*

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.

Copyright © 2021 Stat-Ease, Inc. Do not copy or redistribute in any form.