BASIC DESIGN OF EXPERIMENTS USING CUSTOM DOE PLATFORM

Mastering JMP Webcast
April 24, 2015

Tom Donnelly, PhD, CAP
Sr. Systems Engineer
& Co-insurrectionist
JMP Federal Government Team
AGENDA

• JMP DOE Resources
• Multiple Response Optimization – Trade-Space Analysis
• Review of Classic DOE
• Real-World Experimental Issues – Custom DOE is about
  *Making Designs Fit the Problem –*
  \textit{NOT Making Problems Fit the Designs!}
• All continuous factors
• Continuous, discrete numeric, categorical and blocking factors - with constraints
DOE RESOURCES

• Help in JMP Software
  • Help > Tutorials > DOE Tutorial
  • Help > Books > Design of Experiments Guide > Chapter 3: Building Custom Designs
• Three tabs across home page at www.jmp.com
  • Using JMP > Application Areas > Design of Experiments
  • News and Events > On-Demand Webcasts > Mastering JMP (see more in series)
  • Resources for Users >
    • JMP User Community
    • Learning Library – one-page guides & 1-3 minute videos
    • JMP Blog – search on Brad Jones, Ryan Lekivetz, and Joseph Morgan
    • File Exchange – Add-ins
• www.jmp.com/government-resources DOE Webcasts & Journal Articles
• DOE Books at http://www.jmp.com/software/books.shtml
  • Design and Analysis of Experiments by Douglas Montgomery: A Supplement for Using JMP®
    by Heath Rushing, Andrew Karl and James Wisnowski
  • Optimal Design of Experiments: A Case Study Approach
    by Brad Jones and Peter Goos
• Linked-In: JMP DOE subgroup (505 members) of JMP Professional Network
• YouTube: www.youtube.com/watch?v=R7Dh3RRleGo
WHY USE DOE?

QUicker answers, lower costs, solve bigger problems

• More rapidly answer “what if?” questions
• Do sensitivity and trade-space analysis
• Optimize across multiple responses
• By running efficient subsets of all possible combinations, one can – for the same resources and constraints – solve bigger problems
• By running sequences of designs one can be as cost effective as possible and run no more trials than are needed to get a useful answer
USE JMP TRADE-OFF AND OPTIMIZATION

- Speed: [5.250999, 5.489252]
- Contrast: [0.711209, 0.780651]
- Cosine: [0.196324, 0.321139]
- Desirability: 0.342818

Parameters:
- Sensitizer 1
- Sensitizer 2
- Dye
- Reaction Time
- Desirability
CLASSIC RESPONSE-SURFACE DOE IN A NUTSHELL

Fit requires data from all 3 blocks
Can fit data from blocks 1, 2 or 3
Lack-of-fit

Fit requires data from blocks 1 & 2

Fit requires data from all 3 blocks
**POLYNOMIAL MODELS USED TO CALCULATE SURFACES**

*Block 1*

\[ y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 \]

Run this block 1st to:
- (i) estimate the main effects*
- (ii) use center point to check for curvature.

*Block 2*

\[ y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{23} x_2 x_3 \]

Run this block 2nd to:
- (i) repeat main effects estimate,
- (ii) check if process has shifted,
- (iii) add interaction effects to model if needed.

*Block 3*

\[ y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{23} x_2 x_3 + a_{11} x_1^2 + a_{22} x_2^2 + a_{33} x_3^2 \]

Run this block 3rd to:
- (i) repeat main effects estimate,
- (ii) check if process has shifted,
- (iii) add curvature effects to model if needed.
NUMBER OF UNIQUE TRIALS FOR 3 RESPONSE-SURFACE DESIGNS AND
NUMBER OF QUADRATIC MODEL TERMS VS. NUMBER OF CONTINUOUS FACTORS

If generally running 3, 4 or 5-factor fractional-factorial designs…
1. How many interactions are you not investigating?
2. How many more trials needed to fit curvature?
3. Consider two stages: Definitive Screening + Augmentation
NUMBER OF UNIQUE TRIALS FOR 3 RESPONSE-SURFACE DESIGNS AND NUMBER OF QUADRATIC MODEL TERMS VS. NUMBER OF CONTINUOUS FACTORS

If generally running 3, 4 or 5-factor fractional-factorial designs…
1. How many interactions are you not investigating?
2. How many more trials needed to fit curvature?
3. Consider two stages: Definitive Screening + Augmentation

Unique Trials in Central Composite Design
Unique Trials in Box-Behnken Design
Unique Trials in Custom Design with 6 df for Model Error
Terms in Quadratic Model

36 trial I-optimal response-surface design started as 10-factor DSD and was then augmented with 12 more trials in 6 most important factors.
CLASSIC
DEFINITION OF DOE

PURPOSEFUL CONTROL OF THE INPUTS (FACTORS) IN SUCH A WAY AS TO DEDUCE THEIR RELATIONSHIPS (IF ANY) WITH THE OUTPUT (RESPONSES).
ALTERNATIVE DEFINITION OF DOE

A DOE IS THE SPECIFIC COLLECTION OF TRIALS RUN TO SUPPORT A PROPOSED MODEL.

- If proposed model is simple, e.g. just main effects or 1st order effects ($x_1$, $x_2$, $x_3$, etc.), the design is called a screening DOE
  » Goals include rank factor importance or find a “winner” quickly
  » Used with many (> 6?) factors at start of process characterization

- If the proposed model is more complex, e.g. the model is 2nd order so that it includes two-way interaction terms ($x_1 x_2$, $x_1 x_3$, $x_2 x_3$, etc.) and in the case of continuous factors, squared terms ($x_1^2$, $x_2^2$, $x_3^2$, etc.), the design is called a response-surface DOE
  » Goal is generally to develop a predictive model of the process
  » Used with a few (< 6?) factors after a screening DOE
**QUADRATIC MODEL NOT THAT MUCH BIGGER THAN INTERACTION MODEL**

_fit requires data from all 3 blocks

Can fit data from blocks 1, 2 or 3

Fit requires data from blocks 1 & 2

Fit requires data from all 3 blocks

\[ y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 \]

For k factors there are k main effects

For 3 factors Linear Model has 4 terms

For 6 factors Linear Model has 7 terms

For 10 factors Linear Model has 11 terms

\[ y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 \]

For k factors there are \( k(k-1)/2 \) interaction effects

For 3 factors Interaction Model has 7 terms

For 6 factors Interaction Model has 22 terms

For 10 factors Interaction Model has 56 terms

\[ y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 \]

For k factors there are \( k^2 \) squared effects

For 3 factors Quadratic Model has 10 terms

For 6 factors Quadratic Model has 28 terms

For 10 factors Quadratic Model has 66 terms
REAL-WORLD DESIGN ISSUES

How many experimenters have any of these issues?
Most of these are NOT well treated by classic DOE

• Work with these different kinds of control variables/factors:

  » **Continuous/quantitative?** (Finely adjustable like temperature, speed, force)
  » **Categorical/qualitative?** (Comes in types, like material = rubber, polycarbonate, steel with mixed # of levels; 3 chemical agents, 4 decontaminants, 8 coupon materials…)
  » **Mixture/formulation?** (Blend different amounts of ingredients and the process performance is dependent on the proportions more than on the amounts)
  » **Blocking?** (e.g. “lots” of the same raw materials, multiple “same” machines, samples get processed in “groups” – like “eight in a tray,” run tests over multiple days – i.e. variables for which there shouldn’t be a causal effect)

• Work with **combinations of these four kinds** of variables?

• Certain **combinations cannot be run**? (too costly, unsafe, breaks the process)

• Certain factors are **hard-to-change** (temperature takes a day to stabilize)

• Would like to **add onto existing trials**? (really expensive/time consuming to run)
### Real World Variables

**Responses**
- **Add Response**
- **Remove**
- **Number of Responses**

<table>
<thead>
<tr>
<th>Response Name</th>
<th>Goal</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Maximize</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

**Factors**
- **Add Factor**
- **Remove**
- **Add N Factors**

- **Continuous**
- **Discrete Numeric**
- **Categorical**
- **Blocking**
- **Covariate**
- **Mixture**
- **Constant**
- **Uncontrolled**

**Add Factor** button. Double click on a factor button.
CATEGORICAL FACTORS AND RESPONSES

• **Agents**
  - Agent 1
  - Agent 2
  - Agent 3

• **Decontaminants**
  - Decon 1
  - Decon 2
  - Decon 3
  - Decon 4

• **Materials**
  - Steel
  - Aluminum
  - Glass
  - Polycarbonate
  - CARC (Paint)
  - Viton
  - Kapton
  - Silicone

• **Responses**
  - Pass/Fail
  - Yes/No
  - Not Cracked/Cracked
  - Safe/Caution/Unsafe
  - Not Corroded/Moderately Corroded/Severely Corroded
CONTINUOUS FACTORS AND RESPONSES

• Factors
  ▪ Time
  ▪ Temperature
  ▪ Amount of Agent/Unit Area
  ▪ Wind Speed
  ▪ Humidity

• Responses
  ▪ Evaporation Rate
  ▪ Absorption
  ▪ Adsorption
  ▪ Residual Concentration
Example: Number of Teeth on Bicycle Sprockets

**Evenly Spaced**

<table>
<thead>
<tr>
<th>Teeth</th>
<th>16</th>
<th>19</th>
<th>22</th>
<th>25</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>% Change</td>
<td>18.8%</td>
<td>15.8%</td>
<td>13.6%</td>
<td>12.0%</td>
<td></td>
</tr>
</tbody>
</table>

**Actual Spacing**

<table>
<thead>
<tr>
<th>Teeth</th>
<th>16</th>
<th>18</th>
<th>22</th>
<th>24</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>% Change</td>
<td>12.5%</td>
<td>22.2%</td>
<td>9.1%</td>
<td>16.7%</td>
<td></td>
</tr>
</tbody>
</table>

**Improved Spacing**

<table>
<thead>
<tr>
<th>Teeth</th>
<th>16</th>
<th>18</th>
<th>21</th>
<th>24</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>% Change</td>
<td>12.5%</td>
<td>16.7%</td>
<td>14.3%</td>
<td>16.7%</td>
<td></td>
</tr>
</tbody>
</table>
Only make four sizes of pizza 10”, 12”, 14” & 16”
Mid-point of full range is 13” diameter.

Corresponding areas in sq. in. are: 78, 113, 154 & 201
Mid-point of full range is 123 sq. in. or 12.5” diameter.
Could use either set of four actual sizes in diameter or area as levels of a discrete numeric variable.
**MIXTURE VARIABLES**

**SIMPLE MIXTURE – MAKING SALAD DRESSING**

- Relative *proportions* of factors or components is more important than actual quantity.
- Three liquid components - Oil, Water, and Vinegar.
- 8 oz. in Cruet vs. 4 gal. in Jug:
  - 5 oz. “O” 320 oz. 5/8
  - 1 oz. “W” 64 oz. 1/8
  - 2 oz. “V” 128 oz. 1/4
- To study these mixture components in a DOE use ranges that are proportions:
  - O: 0.500 to 0.750 (½ to ¾)
  - W: 0.000 to 0.250 (0 to ¼)
  - V: 0.125 to 0.375 (⅛ to ⅜)
- Sum of proportions *constrained* to equal 1.

1 = O + W + V so therefore…

W = 1 − (O + V), O = 1 − (V + W), & V = 1 − (O + W)
A design run over 5 days that is sensitive to humidity might SHIFT on Thursday

- But what if because of the rain the tester from days 1, 2, 3 & 5 didn’t make it to work?
- What if that day the power went out briefly? Or, dept. meeting “paused” the work? Or…?

The block variable doesn’t tell you the cause of the effect - just that a shift has been detected among blocks.

- Hoping block variable has no effect. If it does then how can we reliably predict other blocks? If significant it probably means we are missing a factor.
- The only way to be sure that no “unknown” factor has crept into the experiment, is to test for it - and “blocking” your design is not expensive.
- Block variable is a categorical factor having only 1-way effects (no interactions)
TIME AND TEMPERATURE CONSTRAINTS UNCODED

- Time: 10 20
- Temp: 350 450
- Pizza Size: 10, 12, 14 & 16
- Pizza Type
  - Cheese
  - Veggies
  - Meats
- Batch (Size = 5)

- Hi + Hi = “Burnt”
- Lo + Lo = “Not Done”
TIME AND TEMPERATURE CONSTRAINTS UNCODED

- Shorter times means more product produced per hour
- Make most of our money in a few hours each evening

![Graph Builder](Temperature vs. Time)

**MAKE MORE $$$**

**MAKE LESS $**
TIME AND TEMPERATURE
CONSTRAINTS
CODED VS. UNCODED
TIME AND TEMPERATURE CONSTRAINTS CODED

Graph Builder

Temperature vs. Time


TIME AND TEMPERATURE CONSTRAINTS CODED

- $y = mx + b$
- $\text{Temp} = m \times \text{Time} + b$
- $[1] \times \text{Temp} = [-1] \times \text{Time} + [1]$
- $[1] \times \text{Time} + [1] \times \text{Temp} = [1]$

- $[1] \times \text{Time} + [1] \times \text{Temp} \leq [1]$

- $y = mx + b$
- $\text{Temp} = m \times \text{Time} + b$
- $[1] \times \text{Temp} = [-1] \times \text{Time} + [-1]$
- $[1] \times \text{Time} + [1] \times \text{Temp} = [-1]$

- $[1] \times \text{Time} + [1] \times \text{Temp} \geq [-1]$
TIME AND TEMPERATURE
CONSTRAINTS
UNCODED

Graph Builder

Temperature vs. Time


• \( y = mx + b \)
• \( \text{Temp} = m \times \text{Time} + b \)
• \( [1] \times \text{Temp} = [10] \times \text{Time} + [600] \)
• \( [10] \times \text{Time} + [1] \times \text{Temp} = [600] \)
• \( [10] \times \text{Time} + [1] \times \text{Temp} \leq [600] \)

• \( y = mx + b \)
• \( \text{Temp} = m \times \text{Time} + b \)
• \( [1] \times \text{Temp} = [-10] \times \text{Time} + [500] \)
• \( [10] \times \text{Time} + [1] \times \text{Temp} = [500] \)
• \( [10] \times \text{Time} + [1] \times \text{Temp} \geq [500] \)
**TIME AND TEMPERATURE CONSTRAINTS UNCODED**

<table>
<thead>
<tr>
<th>Time</th>
<th>Temp</th>
<th>Constraint Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>450</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>350</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>400</td>
</tr>
</tbody>
</table>

**Bivariate Fit of Time By Temp Constraint Location=Lower**

- Time = 50 - 0.1*Temp
- Time + 0.1*Temp = 50
- 10*Time + 1*Temp = 500
- 10*Time + 1*Temp >= 500

**Bivariate Fit of Time By Temp Constraint Location=Upper**

- Time = 60 - 0.1*Temp
- Time + 0.1*Temp = 60
- 10*Time + 1*Temp = 600
- 10*Time + 1*Temp <= 600

**Linear Fit**

- Time = 50 - 0.1*Temp
- Time = 60 - 0.1*Temp
FINAL DESIGN
SHOWING
CONSTRAINED
REGIONS
FINAL DESIGN SHOWING CONSTRAINED REGIONS

Temperature vs. Time

<table>
<thead>
<tr>
<th>Pizza Type</th>
<th>Cheese</th>
<th>Veggie</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pizza Size:
- • 10
- • 12
- • 14
- • 16

Time vs. Temperature

Meats

Temperature

10.0  15.0  20.0  10.0  15.0  20.0

FINAL DESIGN SHOWING CONSTRAINED REGIONS