There are many attendees today! To avoid disrupting the Voice over Internet Protocol (VoIP) system, I will mute all. Please email Questions to me which I answer after the presentation.

-- Pat

Presented by Pat Whitcomb, Founder Stat-Ease, Inc., Minneapolis, MN pat@statease.com
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE
- Combining categoric factors
- A Couple of Case Studies
  - The Loose Collet
  - A Case to Test Your Metal
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
  Expand your search without sacrificing precision.
  *5 times the volume with no loss in precision!*

- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE

- Combining categoric factors

- A Couple of Case Studies
  - The Loose Collet
  - A Case to Test Your Metal
Four Factor CCD – Case Study #1
alpha = 2.0 (rotatable and spherical)

There are four factors (k=4), but picture only shows three factors.
Prediction standard error of the expected value:

\[
PV(x_0) = \text{var}(\hat{y}_0) = \left(x_0^T \left(X^T X\right)^{-1} x_0\right)s^2
\]

\[
\text{StdErr}(x_0) = s\hat{y}_0 = \sqrt{PV(x_0)}
\]

- \(x_0\) – the location in the design space (i.e. the x coordinates for all model terms).
- \(X\) – the experimental design (i.e. where the runs are in the design space).

In this four factor CCD, the factorial and axial points have:

\[
\text{StdErr} = 3.43744
\]
Four Factor CCD – Case Study #1
alpha = 2.0 (rotatable and spherical)

Slice at: C = 0, D = 0

Slice at: C = +1, D = +1

Yellow border at StdErr = 3.43744
Four Factor CCD – Case Study #1
Maximize Protein with StdErr ≤ 3.43744
Four Factor CCD – Case Study #1
Maximize Protein with StdErr ≤ 3.43744

A: Heating (C / 30 min)
B: pH
C: Redox pot (volt)
D: Na lauryl

Protein (%)
Std Error of Protein

Prediction 86.311
SE Mean 3.43743

A: Heating
-1.579
B: pH
0.234
C: Redox pot
-1.008
D: Na lauryl
-0.660
Protein
86.311
StdErr
3.437

June 2017 Webinar
## Four Factor CCD – Case Study #1
Maximize Protein

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>y</th>
<th>StdErr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heating</td>
<td>pH</td>
<td>Redox Pot</td>
<td>Na lauryl</td>
<td>Protein</td>
<td></td>
</tr>
</tbody>
</table>

**Factor range ±1.00 cube** (too restrictive) *(max SE = 3.43744)*

-1.000  -0.157  -0.870  -0.999  84.696  2.595

**Factor range ±2.00 cube** (too liberal) *(max SE = 13.3764)*

-2.000  0.636  -2.000  -2.000  89.644  10.321

**SE ≤ 3.43744 sphere** (just right) *(max SE = 3.43744)*

-1.579  0.234  -1.008  -0.660  86.311  3.437

Increases volume of search by 393% over ±1 cube
Historical Data – Case Study #2
Not Space Filling for Cube or Sphere

Points projected in two-dimensional planes.
Historical Data – Case Study #2
Limit Search to StdErr ≤ 31.7524

C = -1

C = 0

C = +1

Highest Standard Error at a design point = 31.7524

June 2017 Webinar
Historical Data – Case Study #2
Maximize Tack with StdErr ≤ 31.7524
Historical Data – Case Study #2
Maximum Tack with StdErr ≤ 31.7524

Design-Expert® Software
Factor Coding: Coded
Overlay Plot

Tack
StdErr(Tack)
- Design Points

X1 = A: Mw
X2 = B: Wt% OH

Coded Factor
C: Monomer = 1.000

StdErr(Tack): 31.7524

A: Mw (gm)

B: Wt% OH (wt%)

Tack: 1192.97
X1 -1.00
X2 -0.62
Conclusions: Using Standard Error to Constrain Optimization

Advantages:

- Defines a search area that matches design properties:
  - Spheres for rotatable CCDs.
  - Cubes for face centered CCDs.
  - Irregular shapes for optimal designs and historical data.

- Modifies search area for reduced models and missing data.

June 2017 Webinar
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE
  *Incorporate specifications in your optimization.*
- Combining categoric factors
- A Couple of Case Studies
  - The Loose Collet
  - A Case to Test Your Metal
Process Capability Refresher

$C_{pk}$ statistic

$$C_{pk} = \frac{\text{Minimum}(USL - \mu, \mu - LSL)}{3\sigma}$$

$C_{pk}$ defines the potential capability of the process, with regard to where the process is centered.
**Process Capability Refresher**

**P<sub>pk</sub> statistic**

P<sub>pk</sub> statistic is used to determine how “in control” the process has been over time.

\[
Z_{\text{upper}} = \frac{\text{USL} - \bar{Y}}{3\hat{s}}, \quad Z_{\text{lower}} = \frac{\bar{Y} - \text{LSL}}{3\hat{s}}
\]

P<sub>pk</sub> = minimum(P<sub>pk</sub> upper, P<sub>pk</sub> lower)

P<sub>pk</sub> allows for variation from a centered process.

\( \hat{s} \) is calculated from data gathered over time to capture long term variation.

The difference between C<sub>pk</sub> and P<sub>pk</sub> is the estimate of sigma:

- C<sub>pk</sub> uses a **short** term estimate of sigma
- P<sub>pk</sub> uses a **long** term estimate of sigma
Propogation of Error Tool to Improve $C_{pk}$ & $P_{pk}$
To illustrate the theory, the control factors were used in two steps: first to decrease variation and second to move back on target.

In practice, numerical optimization can be used to simultaneously obtain all the goals.
Propagation of Error
Goal: Minimize Propagated Error (POE)

First order:

\[
\sigma^2 \hat{y} = \sum_i \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma^2_{x_i} + \sigma^2_e \\
\hat{y} = f(x_1, \ldots, x_k)
\]

Second order:

\[
\sigma^2 \hat{y} = \sum_{i=1}^{k} \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma^2_{x_i} + \frac{1}{2} \sum_{i=1}^{k} \left( \frac{\partial^2 f}{\partial x_i^2} \right)^2 \sigma^4_{x_i} + \sum_{i<j} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 \sigma^2_{x_i} \sigma^2_{x_j} + \sigma^2_e \\
\hat{y} = f(x_1, \ldots, x_k) + \frac{1}{2} \sum_{i=1}^{k} \left( \frac{\partial^2 f}{\partial x_i^2} \right) \sigma^2_{x_i}
\]

\[
\sigma^2_e = MS_{\text{residual}} \text{ from the ANOVA and } POE = \sqrt{\sigma^2 \hat{y}}
\]
The difference between $C_{pk}$ and $P_{pk}$ is the estimate of sigma:

- $C_{pk}$ uses a **short** term estimate of sigma
- $P_{pk}$ uses a **long** term estimate of sigma

**Design of Experiments (DOE):**

- The residual MSE ($\sigma_{e}^2$) estimates short variation ($C_{pk}$)
- Propagation of error estimates long term variation ($P_{pk}$ like)

\[
\sigma_Y^2 = \sum_{i=1}^{k} \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + \frac{1}{2} \sum_{i=1}^{k} \left( \frac{\partial^2 f}{\partial x_i^2} \right)^2 \sigma_i^4 + \sum_{i<j} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 \sigma_i^2 \sigma_j^2 + \sigma_e^2
\]
Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE Lathe Case Study

The experimenters run a three factor Box-Behnken design:

The key response is delta, i.e. the deviation of the finished part’s dimension from its nominal value. Delta is measured in mils, 1 mil = 0.001 inches.
Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE

Lathe Case Study

![ANOVA Table]

- **Source**: Model, A-Speed, B-Feed, C-Depth, AB, AC, $A^2$, $C^2$, Residual, Lack of Fit, Pure Error, Cor Total
- **Sum of Squares**: 1.36, 0.037, 0.15, 0.35, 0.27, 0.24, 0.025, 0.27, 0.041, 0.019, 0.022, 1.40
- **df**: 7, 1, 1, 1, 1, 1, 1, 1, 9, 5, 1, 16
- **Mean Square**: 0.19, 0.037, 0.15, 0.35, 0.27, 0.24, 0.025, 0.27, 4.558E-003, 3.740E-003, 5.582E-003
- **F Value**: 42.65, 8.17, 32.42, 77.80, 59.70, 53.21, 5.52, 59.53, 0.67, 0.6689
- **p-value**: < 0.0001, 0.0189, 0.0003, < 0.0001, < 0.0001, < 0.0001, 0.0434, < 0.0001, 0.67, 0.6689

- **Std. Dev.**: 0.068
- **Mean**: -3.535E-003
- **C.V. %**: 1909.78
- **PRESS**: 0.15
- **R-Squared**: 0.9707
- **Adj R-Square**: 0.9480
- **Pred R-Square**: 0.8906
- **Adeq Precision**: 20.957
Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE

Lathe Case Study

Variation of Inputs:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Std Dev: $\sigma_{ii}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A – Speed</td>
<td>5 fpm</td>
</tr>
<tr>
<td>B – Feed</td>
<td>0.00175 ipr</td>
</tr>
<tr>
<td>C – Depth</td>
<td>0.0125 inches</td>
</tr>
</tbody>
</table>

Specifications: $\Delta = 0.000 \pm 0.400$

Goal is to maximize $P_{pk}$:

![Goal Cpk Interface](image)
Employmg $C_{pk}$ (or $P_{pk}$) to optimize your DOE

Lathe Case Study

delta: $C_{pk} = 1.66$

delta (POE): $C_{pk} = 1.11$ (more like $P_{pk}$)
Conclusions: Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE

Advantages:

- Brings specifications into the optimization.
- Can use POE to better represent long term variability.
- In multiple response optimization the various process capabilities can be weighted by importance of response.
- Can explore what if questions, e.g. what if there was better control of various factors; e.g. compare $C_{pk}$ without POE to $C_{pk}$ with POE ($P_{pk}$).
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE
- **Combining categoric factors**
  *Reduce redundancy and number of runs.*
- A Couple of Case Studies
  - The Loose Collet
  - A Case to Test Your Metal
Combining Categoric Factors
CWA detection Case Study

Goal: Prove effectiveness of a remote detection system for a chemical-warfare agent (CWA).

Factors and levels:

A. CWA (Threshold, Objective)
B. Interferent (None, Burning diesel, Burning plastic)
C. Time (Day, Night)
D. Distance (1 km, 3 km, 5 km)
E. Environment (Desert, Tropical, Arctic, Urban, Forest)
F. Season (Summer, Winter)
G. Temperature (High, Low)
H. Humidity (High, Low)
Looking at the last four factors, there are too many combinations (40) and not all of these combinations are meaningful:

E. Environment (Desert, Tropical, Arctic, Urban, Forest)
F. Season (Summer, Winter)
G. Temperature (High, Low)
H. Humidity (High, Low)

Not meaningful: Tropical with low temperature and low humidity
Arctic with high temperature and high humidity

Solution is to combine these four categorical factors into one.
Combining Categoric Factors
CWA detection Case Study

The eight most meaningful combinations:

<table>
<thead>
<tr>
<th>Background</th>
<th>Temperature</th>
<th>Humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desert Winter</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Desert Summer</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Tropical</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Arctic</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Urban Winter</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Urban Summer</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Forest Winter</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Forest Summer</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
Conclusions: Combining Categoric Factors

Advantages:

- Number of runs reduced.
  - Runs reduced by 80% (1,440 to 288) for CWA.

- All the combinations can be run and result in meaningful results.

- Can fit a full model.
  
  Preventing meaningless (or impossible) runs that result in missing data thereby creating an aliased model.
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE
- Combining categoric factors

**A Couple of Case Studies**

- **The Loose Collet**
  Diagnostics plus subject matter knowledge save the day.
- **A Case to Test Your Metal**
1. A computer controlled lathe feeds in bar stock, cuts it, machines the surface and releases a part. The collet holds the part in place as it is being machined. The operator programs the speed (rate of spin) and feed (depth of cut). The operator hand tightens the collet.

2. The response is surface finish, measured on the same one inch section of each part. The higher the reading the rougher the surface. Low readings (a smooth surface) are desirable.

3. The factors studied are:

<table>
<thead>
<tr>
<th>Factor</th>
<th>-1</th>
<th>+1</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Speed</td>
<td>2500</td>
<td>4500</td>
<td>rpm</td>
</tr>
<tr>
<td>B Feed</td>
<td>0.003</td>
<td>0.009</td>
<td>in/rev</td>
</tr>
<tr>
<td>C Collet</td>
<td>Loose</td>
<td>Tight</td>
<td></td>
</tr>
<tr>
<td>D Tool wear</td>
<td>New</td>
<td>after 250 parts</td>
<td></td>
</tr>
</tbody>
</table>

4. The design run was a $2^4$ replicated factorial.

5. Determine what is causing rough surface?
The Loose Collet Analysis Replicated $2^4$

- Error estimates
- Shapiro-Wilk test
  - W-value = 0.792
  - p-value = 0.024

Factors:
- A: Speed
- B: Feed
- C: Collet
- D: Tool wear

Design-Expert® Software
Finish

Half-Normal Plot

<table>
<thead>
<tr>
<th>Half-Normal % Probability</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Standardized Effect]</td>
<td>0.00</td>
<td>35.91</td>
<td>71.81</td>
<td>107.72</td>
<td>143.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:
- Positive Effects
- Negative Effects
The Loose Collet Analysis Replicated $2^4$

ANOVA for selected factorial model

Analysis of variance table [Partial sum of squares - Type III]

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5.400E+005</td>
<td>7</td>
<td>77140.36</td>
<td>15.32</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>A-Speed</td>
<td>56784.50</td>
<td>1</td>
<td>56784.50</td>
<td>11.28</td>
<td>0.0026</td>
</tr>
<tr>
<td>B-Feed</td>
<td>21218.00</td>
<td>1</td>
<td>21218.00</td>
<td>4.21</td>
<td>0.0511</td>
</tr>
<tr>
<td>C-Collet</td>
<td>1.650E+005</td>
<td>1</td>
<td>1.650E+005</td>
<td>32.78</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>55444.50</td>
<td>1</td>
<td>55444.50</td>
<td>11.01</td>
<td>0.0029</td>
</tr>
<tr>
<td>AC</td>
<td>44253.13</td>
<td>1</td>
<td>44253.13</td>
<td>8.79</td>
<td>0.0067</td>
</tr>
<tr>
<td>BC</td>
<td>1.423E+005</td>
<td>1</td>
<td>1.423E+005</td>
<td>28.27</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>ABC</td>
<td>54946.13</td>
<td>1</td>
<td>54946.13</td>
<td>10.91</td>
<td>0.0030</td>
</tr>
<tr>
<td>Residual</td>
<td>1.208E+005</td>
<td>24</td>
<td>5034.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit</td>
<td>6763.00</td>
<td>8</td>
<td>845.38</td>
<td>0.12</td>
<td>0.9976</td>
</tr>
<tr>
<td>Pure Error</td>
<td>1.141E+005</td>
<td>16</td>
<td>7128.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cor Total</td>
<td>6.608E+005</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Std. Dev. 70.95
Mean 203.63
C.V. % 34.85
PRESS 2.148E+005
The Loose Collet
Diagnostic Plots Replicated 2

Residual Plots
(runs 1, 4 & 20 are highlighted)

Normal Plot of Residuals
Residuals vs. Predicted
Residuals vs. Run

Externally Studentized Residuals
Normal % Probability
Normal Plot of Residuals
-6.00 -4.00 -2.00 0.00 2.00 4.00 6.00
1
5
10
20
30
50
70
80
90
95
99

Predicted
Externally Studentized Residuals
Residuals vs. predicted
-6.00 -4.00 -2.00 0.00 2.00 4.00 6.00
0 100 200 300 400 500
3.58606
-3.58606
0

Run Number
Externally Studentized Residuals
Residuals vs. Run
-6.00 -4.00 -2.00 0.00 2.00 4.00 6.00
1 6 11 16 21 26 31
3.58606
-3.58606
0
The residuals from runs 1, 4 and 20 seem unusually large:

<table>
<thead>
<tr>
<th>Std</th>
<th>Run</th>
<th>A: Speed (rpm)</th>
<th>B: Feed (in/rev)</th>
<th>C: Collet</th>
<th>D: Tool wear</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>2500</td>
<td>0.003</td>
<td>Loose</td>
<td>New</td>
<td>101</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>2500</td>
<td>0.003</td>
<td>Loose</td>
<td>New</td>
<td>130</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4500</td>
<td>0.003</td>
<td>Loose</td>
<td>New</td>
<td>273</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4500</td>
<td>0.003</td>
<td>Loose</td>
<td>New</td>
<td>691</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>2500</td>
<td>0.009</td>
<td>Loose</td>
<td>New</td>
<td>253</td>
</tr>
<tr>
<td>16</td>
<td>27</td>
<td>4500</td>
<td>0.009</td>
<td>Tight</td>
<td>New</td>
<td>245</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>2500</td>
<td>0.003</td>
<td>Loose</td>
<td>After 250</td>
<td>298</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>2500</td>
<td>0.003</td>
<td>Loose</td>
<td>After 250</td>
<td>87</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>4500</td>
<td>0.009</td>
<td>Tight</td>
<td>After 250</td>
<td>216</td>
</tr>
<tr>
<td>32</td>
<td>7</td>
<td>4500</td>
<td>0.009</td>
<td>Tight</td>
<td>After 250</td>
<td>217</td>
</tr>
</tbody>
</table>

These three runs occur when at slow feed and a loose collet. When measuring the surface finish, it was noted that this treatment combination produced an unusual finish. The surface profile looked like a sine wave. The problem is clear; the bar oscillated and caused an uneven cut pattern.
The Loose Collet
Graph Columns

What we’ve learned so far:

- A loose collet is consistently bad.
- Never use a loose collet in production.

Next steps:

- Ignore loose collet data.
- Reanalyze the remaining data (a replicated $2^3$ factorial).
The Loose Collet Analysis w/o Loose Collet

Half-Normal Plot

<table>
<thead>
<tr>
<th>Standardized Effect</th>
<th>Half-Normal % Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td>46.22</td>
<td>10</td>
</tr>
<tr>
<td>92.44</td>
<td>20</td>
</tr>
<tr>
<td>138.66</td>
<td>30</td>
</tr>
<tr>
<td>184.88</td>
<td>50</td>
</tr>
</tbody>
</table>

ANNOVA for selected factorial model

Analysis of variance table [Partial sum of squares - Type III]

<table>
<thead>
<tr>
<th>Source</th>
<th>Squares</th>
<th>df</th>
<th>Square</th>
<th>Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.367E+005</td>
<td>1</td>
<td>1.367E+005</td>
<td>1479.86</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>B-Feed</td>
<td>1.367E+005</td>
<td>1</td>
<td>1.367E+005</td>
<td>1479.86</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>1293.37</td>
<td>14</td>
<td>92.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit</td>
<td>1061.87</td>
<td>6</td>
<td>176.98</td>
<td>6.12</td>
<td>0.0113</td>
</tr>
<tr>
<td>Pure Error</td>
<td>231.50</td>
<td>8</td>
<td>28.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cor Total</td>
<td>1.380E+005</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Std. Dev.  | 9.61 | R-Squared | 0.9906 |
| Mean       | 131.81 | Adj R-Squarec | 0.9900 |
| C.V. %     | 7.29 | Pred R-Square | 0.9878 |
| PRESS      | 1689.31 | Adeq Precision | 54.403 |

June 2017 Webinar
Take Away:

- Feed (depth of cut) is the only factor that affects surface finish when you use a tight collet!
- A simple solution that could have been complicated if we had:
  - Not paid attention to the residual plots.
  - Not used our subject matter knowledge to ignore loose collets!
Practical DOE “Tricks of the Trade”

- Using standard error to constrain optimization
- Employing $C_{pk}$ (or $P_{pk}$) to optimize your DOE
- Combining categoric factors

A Couple of Case Studies

- The Loose Collet
- A Case to Test Your Metal
  Trick to detect outlier and investigation save the day.
A customer of Stat-Ease examined his aluminum casting process using a fractional factorial design. He studied five factors in a $2^5-1$ factorial design:

A) Hot oil temperature
B) Trip in mm
C) Molten aluminum temperature
D) Fast shot velocity
E) Dwell time

The fraction defective out of 100 parts, was recorded for each design point. To his dismay, none of the factors seemed to make any difference.

* From Stat-Ease client files: Dave DeVowe, Tool Products
Look at the Response Data

<table>
<thead>
<tr>
<th>Select</th>
<th>Std</th>
<th>Run</th>
<th>Factor 1 A: Hot Oil Degrees F</th>
<th>Factor 2 B: Trip mm</th>
<th>Factor 3 C: Metal Degrees F</th>
<th>Factor 4 D: Fast Shot mm</th>
<th>Factor 5 E: Dwell sec</th>
<th>Response 1 Defects Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1</td>
<td>350</td>
<td>390</td>
<td>1300</td>
<td>2.20</td>
<td>5.50</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>350</td>
<td>410</td>
<td>1300</td>
<td>2.20</td>
<td>3.50</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>450</td>
<td>390</td>
<td>1300</td>
<td>1.60</td>
<td>5.50</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>450</td>
<td>390</td>
<td>1300</td>
<td>2.20</td>
<td>3.50</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>350</td>
<td>410</td>
<td>1260</td>
<td>2.20</td>
<td>5.50</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>450</td>
<td>410</td>
<td>1300</td>
<td>2.20</td>
<td>5.50</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>450</td>
<td>410</td>
<td>1300</td>
<td>1.60</td>
<td>5.50</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>450</td>
<td>410</td>
<td>1300</td>
<td>1.60</td>
<td>3.50</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>350</td>
<td>390</td>
<td>1260</td>
<td>1.60</td>
<td>5.50</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>350</td>
<td>390</td>
<td>1260</td>
<td>2.20</td>
<td>3.50</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>450</td>
<td>410</td>
<td>1260</td>
<td>2.20</td>
<td>3.50</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>450</td>
<td>410</td>
<td>1260</td>
<td>1.60</td>
<td>3.50</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>350</td>
<td>390</td>
<td>1260</td>
<td>1.60</td>
<td>5.50</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>350</td>
<td>410</td>
<td>1260</td>
<td>1.60</td>
<td>3.50</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>450</td>
<td>390</td>
<td>1260</td>
<td>2.20</td>
<td>3.50</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>350</td>
<td>390</td>
<td>1300</td>
<td>1.60</td>
<td>3.50</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

Note: Lots of variation in the response!
Notice all the effects fall in a line!

### Half-Normal Plot

Select significant terms - see Tips

- **Positive Effects**
- **Negative Effects**

**Design-Expert® Software**

ArcSin(Sqrt(Defects))

**Shapiro-Wilk test**

- W-value = 0.974
- p-value = 0.909

- **A**: Hot Oil
- **B**: Trip
- **C**: Metal
- **D**: Fast Shot
- **E**: Dwell

Line defaults to fit smallest 50% of effects
Lots of variation in the responses, but no effects!

Procedure to detect a possible outlier:

1. Pick a couple of the most extreme effects on the half normal plot.
2. Look for a potential outlier.
4. Investigate the suspected outlier.
5. “Ignore” the outlier and reanalyze the design.
1. Pick a couple of the most extreme effects on the half normal plot.
2. Look for a potential outlier.

Design-Expert® Software
ArcSin( Sqrt(Defects) )

Color points by value of ArcSin( Sqrt(Defects) ):
- 1.571
- 0.247

Std # 1 Run # 9
X: -5.997
Y: 3.1

Normal Plot of Residuals

Externally Studentized Residuals
2. Look for a potential outlier.
4. Investigate the suspected outlier. 
   *(The operator confirmed problems with run 9 (std 1))*

5. “Ignore” the outlier and reanalyze the design.

Design-Expert® Software
ArcSin( Sqrt(Defects) )

Shapiro-Wilk test
W-value = 0.941
p-value = 0.529
A: Hot Oil
B: Trip
C: Metal
D: Fast Shot
E: Dwell

Positive Effects
Negative Effects
A Case to Test Your Metal Analysis – Second Pass

Do not use a 390 “Trip” with a 1.60 “Fast Shot”:

Design-Expert® Software
Factor Coding: Actual
Original Scale
Defects (Fraction)

X1 = B: Trip
X2 = D: Fast Shot

Actual Factors
A: Hot Oil = 400
C: Metal = 1280
E: Dwell = 4.50

- D: 1.60
- D+: 2.20
References


Stat-Ease Training: Sharpen Up via Computer-Intensive Workshops

Shari Kraber, Workshop Manager & Master Statistician
shari@statease.com

PreDOE
Web-Based (optional)
Stat-Ease Academy

Basic Statistics for Design of Experiments

Designed Experiments for Pharma, Life Sciences, Assay Optimization, Food Science

Plus more classes here!

www.statease.com/training/stat-ease-academy.html

June 2017 Webinar
Practical DOE – “Tricks of the Trade”

Reminder, this presentation is posted at:

www.statease.com/training/webinar.html

If you have additional questions email them to:

pat@statease.com

Thank you for joining me today!