Control of Manufacturing Processes

Subject 2.830/6.780/ESD.63
Spring 2008
Lecture #15

Response Surface Modeling and Process Optimization

April 8, 2008
Outline

• Last Time
  – Fractional Factorial Designs
  – Aliasing Patterns
  – Implications for Model Construction

• Today
  – Response Surface Modeling (RSM)
    • Regression analysis, confidence intervals
  – Process Optimization using DOE and RSM

Reading: May & Spanos, Ch. 8.1 – 8.3
Regression Fundamentals

- Use least square error as measure of goodness to estimate coefficients in a model
- One parameter model:
  - Model form
  - Squared error
  - Estimation using normal equations
  - Estimate of experimental error
  - Precision of estimate: variance in $b$
  - Confidence interval for $\beta$
  - Analysis of variance: significance of $b$
  - Lack of fit vs. pure error
- Polynomial regression
Measures of Model Goodness – $R^2$

• Goodness of fit – $R^2$
  – Question considered: how much better does the model do than just using the grand average?

$$R^2 = \frac{SS_T}{SS_D}$$

  – Think of this as the fraction of squared deviations (from the grand average) in the data which is captured by the model

• Adjusted $R^2$
  – For “fair” comparison between models with different numbers of coefficients, an alternative is often used

$$R^2_{adj} = 1 - \frac{SS_R/\nu_R}{SS_D/\nu_D} = 1 - \frac{s_R^2}{s_D^2}$$

  – Think of this as $(1 – \text{variance remaining in the residual})$.

Recall $\nu_R = \nu_D - \nu_T$
Least Squares Regression

- We use **least-squares** to estimate coefficients in typical regression models
- \[ y_i = \beta x_i + \epsilon_i, \quad i = 1, 2, \ldots, n; \quad \epsilon_i \sim N(0, \sigma^2) \]
- \[ \hat{y}_i = bx_i \]
- Goal is to estimate \( \beta \) with “best” \( b \)
- How define “best”?
  - That \( b \) which minimizes sum of squared error between prediction and data
  \[
  SS(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}x_i)^2
  \]
  - The residual sum of squares (for the best estimate) is
  \[
  SS_{\text{min}} = \sum_{i=1}^{n} (y_i - bx_i)^2 = SS_R
  \]
Least Squares Regression, cont.

• Least squares estimation via normal equations
  – For linear problems, we need not calculate $SS(\beta)$; rather, direct solution for $b$ is possible
  – Recognize that vector of residuals will be normal to vector of $x$ values at the least squares estimate

\[
\begin{align*}
\sum (y - \hat{y})x &= 0 \\
\sum (y - bx)x &= 0 \\
\sum xy &= \sum bx^2 \\
\Rightarrow b &= \frac{\sum xy}{\sum x^2}
\end{align*}
\]

• Estimate of experimental error
  – Assuming model structure is adequate, estimate $s^2$ of $\sigma^2$ can be obtained:

\[
s^2 = \frac{SS_R}{n-1}
\]
Precision of Estimate: Variance in $b$

- We can calculate the variance in our estimate of the slope, $b$:

$$b = \frac{\sum xy}{\sum x^2} \quad \Rightarrow \quad \hat{V}(b) = \frac{s^2}{\sum x_i^2}$$

$$\text{s.e.}(b) = \sqrt{\hat{V}(b)}$$

$$b \pm \text{s.e.}(b)$$

- Why?

$$b = \frac{x_1}{x_2} \cdot y_1 + \frac{x_2}{x_2} \cdot y_2 + \cdots + \frac{x_n}{x_2} \cdot y_n$$

$$= a_1 y_1 + a_2 y_2 + \cdots + a_n y_n$$

$$V(b) = (a_1^2 + a_2^2 + \cdots + a_n^2) \sigma^2$$

$$= \left[ \left( \frac{x_1}{x_2} \right)^2 + \cdots + \left( \frac{x_n}{x_2} \right)^2 \right] \sigma^2$$

$$= \frac{\sum x^2}{(\sum x^2)^2} \sigma^2$$

$$= \frac{\sigma^2}{\sum x^2}$$
Confidence Interval for $\beta$

- Once we have the standard error in $b$, we can calculate confidence intervals to some desired $(1-\alpha)100\%$ level of confidence

\[
\frac{b-\beta}{\text{s.e.}(b)} \sim t \quad \Rightarrow \quad \beta = b \pm t_{\alpha/2} \cdot \text{s.e.}(b)
\]

- Analysis of variance
  - Test hypothesis: $H_0: \beta = b = 0$
  - If confidence interval for $\beta$ includes 0, then $\beta$ not significant
    \[
    \sum y_i^2 = \sum \hat{y}_i^2 + \sum (y_i - \hat{y}_i)^2
    \]
    \[
    n = p + n - p
    \]
  - Degrees of freedom (need in order to use t distribution)
    \[
    p = \# \text{ parameters estimated by least squares}
    \]
Example Regression

<table>
<thead>
<tr>
<th>Age</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>6.16</td>
</tr>
<tr>
<td>22</td>
<td>9.88</td>
</tr>
<tr>
<td>35</td>
<td>14.35</td>
</tr>
<tr>
<td>40</td>
<td>24.06</td>
</tr>
<tr>
<td>57</td>
<td>30.34</td>
</tr>
<tr>
<td>73</td>
<td>32.17</td>
</tr>
<tr>
<td>78</td>
<td>42.18</td>
</tr>
<tr>
<td>87</td>
<td>43.23</td>
</tr>
<tr>
<td>98</td>
<td>48.76</td>
</tr>
</tbody>
</table>

**Whole Model**

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>8836.6440</td>
<td>8836.64</td>
<td>1093.146</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>64.6695</td>
<td>8.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>9</td>
<td>8901.3135</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tested against reduced model: Y=0

**Parameter Estimates**

| Term       | Estimate | Std Error | t Ratio | Prob>|t| |
|------------|----------|-----------|---------|------|------|
| Intercept  | Zeroed   | 0         | 0       | .    | .    |
| age        |          | 0.500983  | 0.015152| 33.06| <.0001|

**Effect Tests**

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratic</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1</td>
<td>1</td>
<td>8836.6440</td>
<td>1093.146</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

- Note that this simple model assumes an intercept of zero – model must go through origin
- We can relax this requirement
Lack of Fit Error vs. Pure Error

• Sometimes we have replicated data
  – E.g. multiple runs at same x values in a designed experiment

• We can decompose the residual error contributions

\[ SS_R = SS_L + SS_E \]

Where
\[ SS_R = \text{residual sum of squares error} \]
\[ SS_L = \text{lack of fit squared error} \]
\[ SS_E = \text{pure replicate error} \]

• This allows us to TEST for lack of fit
  – By “lack of fit” we mean evidence that the linear model form is inadequate

\[ \frac{s^2_L}{s^2_E} \sim F_{\nu_L,\nu_E} \]
Regression: Mean Centered Models

- Model form \( y = \alpha + \beta(x - \bar{x}) \)
- Estimate by \( \hat{y} = a + b(x - \bar{x}), \quad (y_i - \hat{y}_i) \sim N(0, \sigma^2) \)

Minimize \( SS_R = \sum (y_i - \hat{y}_i)^2 \) to estimate \( \alpha \) and \( \beta \)

\[
\begin{align*}
a &= \bar{y} \\
E(a) &= \alpha \\
\text{Var}(a) &= \text{Var} \left[ \frac{\sum y_i}{n} \right] = \frac{\sigma^2}{n}
\end{align*}
\]

\[
\begin{align*}
b &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\
E(b) &= \beta \\
\text{Var}(b) &= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}
\end{align*}
\]
Regression: Mean Centered Models

- **Confidence Intervals**

\[
\hat{y}_i = \bar{y} + b(x_i - \bar{x})
\]

\[
\text{Var}(\hat{y}_i) = \text{Var}(\bar{y}) + (x_i - \bar{x})^2 \text{Var}(b)
\]

\[
= \frac{s^2}{n} + \frac{s^2(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2} = s^2_{\hat{y}_i}
\]

- **Our confidence interval on output \( y \) widens as we get further from the center of our data!**

\[
\hat{y}_i \pm t_{\alpha/2} \cdot s_{\hat{y}_i}
\]
Polynomial Regression

• We may believe that a higher order model structure applies. Polynomial forms are also linear in the coefficients and can be fit with least squares

\[ \eta = \beta_0 + \beta_1 x + \beta_2 x^2 \]

Curvature included through \( x^2 \) term

• Example: Growth rate data
Regression Example: Growth Rate Data

<table>
<thead>
<tr>
<th>Observation Number</th>
<th>Amount of Supplement (grams) x</th>
<th>Growth Rate (coded units) y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>85</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>91</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>35</td>
<td>65</td>
</tr>
</tbody>
</table>

Growth rate data

- Replicate data provides opportunity to check for lack of fit
Growth Rate – First Order Model

- Mean significant, but linear term not
- Clear evidence of lack of fit

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$S_M = 67,428.6 { \text{mean: 67,404.1} } \text{ extra for linear: 24.5}$</td>
<td>$2 { 1 }$</td>
<td>67,404.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24.5</td>
</tr>
<tr>
<td>Residual</td>
<td>$S_R = 686.4 { S_L = 659.40 } \text{ pure error: 27.0}$</td>
<td>$8 { 4 }$</td>
<td>164.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.75</td>
</tr>
<tr>
<td>Total</td>
<td>$S_T = 68,115.0$</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*Analysis of variance for growth rate data: Straight line model*
Growth Rate – Second Order Model

- No evidence of lack of fit
- Quadratic term significant

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SM = 68,071.8</td>
<td>{ mean 67,404.1, extra for linear 24.5, extra for quadratic 643.2 }</td>
<td>67,404.1</td>
</tr>
<tr>
<td>Model</td>
<td>SR = 43.2</td>
<td>7 3 { 1 } { 1 } { 1 }</td>
<td>24.5 643.2</td>
</tr>
<tr>
<td>Residual</td>
<td>SE = 27.0</td>
<td>10 3 { 4 }</td>
<td>5.40 6.75</td>
</tr>
<tr>
<td>Total</td>
<td>ST = 68,115.0</td>
<td>{ 1 } { 1 } { 1 }</td>
<td>ratio = 0.80</td>
</tr>
</tbody>
</table>

*Analysis of variance for growth rate data: Quadratic model*
Polynomial Regression In Excel

- Create additional input columns for each input
- Use “Data Analysis” and “Regression” tool

<table>
<thead>
<tr>
<th>x</th>
<th>x^2</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>15</td>
<td>225</td>
<td>85</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>90</td>
</tr>
<tr>
<td>25</td>
<td>625</td>
<td>91</td>
</tr>
<tr>
<td>30</td>
<td>900</td>
<td>75</td>
</tr>
<tr>
<td>35</td>
<td>1225</td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R: 0.968</td>
</tr>
<tr>
<td>R Square: 0.936</td>
</tr>
<tr>
<td>Adjusted R Square: 0.918</td>
</tr>
<tr>
<td>Standard Error: 2.541</td>
</tr>
<tr>
<td>Observations: 10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>Regression</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>35.657</td>
<td>5.618</td>
<td>6.347</td>
<td>22.373</td>
<td>48.942</td>
</tr>
<tr>
<td>x</td>
<td>5.263</td>
<td>0.558</td>
<td>9.431</td>
<td>3.943</td>
<td>6.582</td>
</tr>
<tr>
<td>x^2</td>
<td>-0.128</td>
<td>0.013</td>
<td>-9.966</td>
<td>-0.158</td>
<td>-0.097</td>
</tr>
</tbody>
</table>

2.830J/6.780J/ESD.63J
Polynomial Regression

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Square</th>
<th>Mean Square</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>665.70617</td>
<td>332.853</td>
<td>51.5551</td>
</tr>
<tr>
<td>Error</td>
<td>7</td>
<td>45.19383</td>
<td>6.456</td>
<td>Prob &gt; F</td>
</tr>
<tr>
<td>C. Total</td>
<td>9</td>
<td>710.90000</td>
<td></td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

**Lack Of Fit**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Square</th>
<th>Mean Square</th>
<th>F Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack Of Fit</td>
<td>3</td>
<td>18.193829</td>
<td>6.0646</td>
<td>0.8985</td>
</tr>
<tr>
<td>Pure Error</td>
<td>4</td>
<td>27.000000</td>
<td>6.7500</td>
<td>Prob &gt; F</td>
</tr>
<tr>
<td>Total Error</td>
<td>7</td>
<td>45.193829</td>
<td></td>
<td>0.5157</td>
</tr>
</tbody>
</table>

**Summary of Fit**

- RSquare: 0.936427
- RSquare Adj: 0.918264
- Root Mean Sq Error: 2.540917
- Mean of Response: 82.1
- Observations (or Sum Wgts): 10

**Parameter Estimates**

| Term   | Estimate | Std Error | t Ratio | Prob>|t| |
|--------|----------|-----------|---------|-----|
| Intercept | 35.657437 | 5.617927 | 6.35   | 0.0004 |
| x      | 5.2628956 | 0.558022 | 9.43   | <.0001 |
| x*x    | -0.127674 | 0.012811 | -9.97  | <.0001 |

**Effect Tests**

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1</td>
<td>1</td>
<td>574.28553</td>
<td>88.9502</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x*x</td>
<td>1</td>
<td>1</td>
<td>641.20451</td>
<td>99.3151</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

- • Generated using JMP package

2.830J/6.780J/ESD.63J
Outline

• Response Surface Modeling (RSM)
  – Regression analysis, confidence intervals

• Process Optimization using DOE and RSM
  – Off-line/iterative
  – On-live/evolutionary
Process Optimization

• Multiple Goals in “Optimal” Process Output
  – Target mean for output(s) $Y$
  – Small variation/sensitivity

$$
\Delta Y = \frac{\partial Y}{\partial \alpha} \Delta \alpha + \frac{\partial Y}{\partial u} \Delta u
$$

• Can Combine in an Objective Function “$J$”
  – Minimize or Maximize, e.g. $\min_x J$ $\max_x J$
  – Such that $J = J(\text{factors})$; might include $J(x); J(\alpha)$

• Adjust $J$ via factors with constraints
Methods for Optimization

• Analytical Solutions
  – $\frac{\partial y}{\partial x} = 0$

• Gradient Searches
  – Hill climbing (steepest ascent/descent)
  – Local min or max problem
  – Excel solver given a convex function

• Offline vs. Online
Basic Optimization Problem

\[ y^0 = J^0 \]

\[ y \text{ (or } J) \]

\[ x^0 \]

\[ x \]
3D Problem

![3D Problem Diagram]

2.830J/6.780J/ESD.63J
Analytical

\[ \frac{\partial y(x)}{\partial x} = 0 \]

- Need Accurate \( y(x) \)
  - Analytical Model
  - Dense \( x \) increments in experiment
- Difficult with Sparse Experiments
  - Easy to missing optimum
Sparse Data Procedure – Iterative Experiments/Model Construction

- Linear models with small increments
- Move along desired gradient
- Near zero slope change to quadratic model
Extension to 3D
Linear Model Gradient Following

\[ \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \]
Steepest Descent

\[ \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \]

\[ g_{x_1} = \frac{\partial y}{\partial x_1} = \beta_1 + \beta_{12} x_2 \]

\[ g_{x_2} = \frac{\partial y}{\partial x_2} = \beta_2 + \beta_{12} x_1 \]

Make changes in \(x_1\) and \(x_2\) along \(G\)

\[ \Delta x_2 = \frac{g_{x_1}}{g_{x_2}} \Delta x_1 \]
Various Surfaces

2.830J/6.780J/ESD.63J
A Procedure for DOE/Optimization

• Study Physics of Process
  – Define important inputs
  – Intuition about model
  – Limits on inputs

• DOE
  – Factor screening experiments
  – Further DOE as needed
  – RSM Construction

• Define Optimization/Penalty Function
  – $J = f(x)$
    \[
    \max_J \quad \min_J \\
    x \quad \quad \quad \quad x
    \]

For us, $x = \mu$ or $\alpha$
(1) DOE Procedure

- Identify model (linear, quadratic, terms to include)
- Define inputs and ranges
- Identify “noise” parameters to vary if possible ($\Delta\alpha$’s)
- Perform experiment
  - Appropriate order
    - randomization
    - blocking against nuisance or confounding effects
(2) RSM Procedure

• Solve for $\beta$’s
• Apply ANOVA
  – Data significant?
  – Terms significant?
  – Lack of Fit significant?
• Drop Insignificant Terms
• Add Higher Order Terms as needed
(3) Optimization Procedure

- Define Optimization/Penalty Function
- Search for Optimum
  - Analytically
  - Piecewise
  - Continuously/evolutionary
- Confirm Optimum
Confirming Experiments

- Checking intermediate points
- Data only at corners
- Test at interior point
- Evaluate error
- Consider Central Composite?

- Rechecking the “optimum”
Optimization Confirmation Procedure

• Find optimum value \( x^* \)

• Perform confirming experiment
  – Test model at \( x^* \)
  – Evaluate error with respect to model
  – Test hypothesis that \( y(x^*) = \hat{y}(x^*) \)

• If hypothesis fails
  – Consider new ranges for inputs
  – Consider higher order model as needed
  – Boundary may be optimum!
Experimental Optimization

• WHY NOT JUST PICK BEST POINT?

• Why not optimize on-line?
  – Skip the Modeling Step?

• Adaptive Methods
  – Learn how best to model as you go
    • e.g. Adaptive OFACT
On-Line Optimization

- Perform $2^k$ Experiment
- Calculate Gradient
- Re-center $2^k$ Experiment About Maximum Corner
- Repeat
- Near Maximum?
  - Should detect quadratic error
  - Do quadratic fit near maximum point
    - Central Composite is good choice here
  - Can also scale and rotate about principal axes
Continuous Optimization: EVOP

- Evolutionary Operation

- Pick “best” $y_i$
- Re-center process
- Do again
Summary

• Response Surface Modeling (RSM)
  – Regression analysis, confidence intervals

• Process Optimization using DOE and RSM
  – Off-line/iterative
  – On-live/evolutionary

• Next Time:
  – Process Robustness
  – Variation Modeling
  – Taguchi Approach