• Design of experiments (DOE) is a powerful tool for product and process improvement.

• JMP is well known as one of the leading software products for the design and analysis of experiments.

• JMP Pro extends modeling capabilities of JMP to more sophisticated data mining models, but is really so much more than that!

• Generalized Regression is a JMP Pro platform for linear models that has powerful tools for analyzing observational data as well as DOE data!
• Historically, analysis of DOEs tends to reflect the computational technology of the time:

  • Orthogonal designs -> Easy to compute coefficients.

  • Transformations -> Stabilize variance with a single transformation of the responses (log, sqrt, inverse).

  • VIFs as a measure of multicollinear inputs.

  • “Manual backward selection” workflow -> Fit full model, remove terms with large p-values, refit model and repeat.
• As computational power and user interfaces improve, better and more direct approaches are possible:
  
  • Model selection should be an integral component of the analysis.
  
  • The entire modeling process should be highly visual and interactive.
  
  • Models using non-normal distributions are a better way to handle variance heterogeneity than transforming the response and then running a least squares analysis.
  
  • Tradeoff analysis of different models should be quick and easy using instantly responsive visual tools.
Generalized Regression (GenReg) in JMP Pro 12 is a game changer in how DOEs are analyzed:

- One-stop shopping for analyzing DOEs since model selection and extraction of useful information (Profilers, diagnostics, multiple comparisons) from the model are all located in the same place.
- Like having stepwise, least squares, and generalized linear models and logistic all in the same place, but is really so much more!
- Learning a little GenReg goes a long way:
  - **Common interface for many different models!**
  - Least Sq., logistic, Poisson, quantile regression, etc.
  - Cox PH, censored responses coming in JMP Pro 13
TODAY’S GOALS

- Use case studies to demonstrate a *fully modern model selection-based approach that emphasizes interactive tools* to assess the practical importance of experimental factors.

- Traditional approaches start with the “full” model and possibly prune the model by removing statistically insignificant factors.

- We propose what amounts to a hybrid approach to analyzing DOEs that is part algorithmic, part interactive:
  
  1) Identify a set of plausible candidate models.
  2) Use interactive tools in JMP along with your subject matter knowledge to choose the best one.
• Demonstrate how to leverage the Solution Path plot as a way to interpret the data and explore different models.

• Use Variable Importance in JMP Profiler to assess which factors are the most important predictors of the response.
From “Statistics For Experimenters” by Box, Hunter, and Hunter.

Five factor, 32 run full factorial to optimize the percent reacted in a nuclear reactor.
• Right-click on the “Model” script, this brings up Fit Model, switch the personality to Generalized Regression, and click Run.
For well-designed experiments like this one, I recommend using Forward Selection and the AICc to find the recommended set of factors and interactions.
The Solution Path (SP) is really two plots:

- Left: Plot of the model coefficients per step in the algorithm.
- Right: Plots the AICc model-selection criteria by step.
- The red lines correspond to the "Goldilocks" model that optimizes goodness of fit and model complexity.
The Solution Path makes it easy to see what the model fitting/selection algorithm is doing:

1) Compute p-values for all the effects eligible to enter the model while respecting the Effect Heredity Rule.

2) Add the term with the smallest p-value to the model, fit the new model, and calculate the models AICc (or other model-selection criteria generally).

3) If there are no more terms that can be added, then STOP, otherwise GOTO (1).
• The goal in DOE analysis is to find the model (set of main effects and polynomial terms) that just the terms that are predictive of the response and without the ones that do not drive the response. The we use that model for prediction, optimization, product improvement, etc.

• We can always improve the fit (reduce SSE) by adding more terms to the model, regardless of whether the term is actually related to the response or not.

• If adding terms always improves the model, how do we know when to stop adding terms to the model? How do we decide which model is the best, or which ones are the good ones?
The model ultimately used balances several considerations:

1. Does that model fit the data well? (goodness of fit)
2. Does the model have too many terms (model complexity)
3. Does the model make sense relative to our subject matter expertise and experience?
4. What is the goal of the current experiment, factor screening or prediction?

Model selection criteria like the AICc and BIC offer guidance on what the data says about the tradeoff of model complexity vs. goodness of fit. (1. and 2.)

The practitioner uses 3. to decide add terms to the model via forcing or choosing a particular model in the path.

In screening one might tolerate more Type I errors, adding more terms from the solution path. Prediction one may be pickier. Again, model selection criteria offer guidance.
The AICc estimates the tradeoff between goodness of fit and model complexity. Experience has shown us that the AICc is a good guide to choosing models via selecting models with low AICc values.

\[
\text{AICc} = n \log(\text{SSE}/n) + 2p + 2p(p+1)/(n-p-1) + \text{constant}.
\]

As Forward Selection adds terms to the model, the SSE goes down (decreasing AICc), but increasing p serves to increase the AICc.

“Model Selection and Multimodel Inference” by Burnham and Anderson is an excellent book on how to use the AICc.
• Usually, early in FS the AICc decreases, reaches its lowest point, and then climbs up as FS ends at the full model with all the possible terms in it.

• Models left of the red line are “too simple,” models to the right are “too complicated.”

• The red line is the “Goldilocks” model and has the “best” tradeoff of goodness of fit to model complexity.

• “Green Zone” models are strongly consistent with the best model. Green Zone = Best AICc+4.

• “Yellow Zone” models are strongly consistent with the best model. Yellow Zone = Best AICc+10.
• The BIC is another popular criteria which is used similarly to the AICc.

• BIC = $n \log(\text{SSE}/n) + p \log(n) + \text{constant}$.

• BIC tends to select models with more terms than the AICc with small datasets. I use BIC over AIC sometimes in screening situations.
USING THE SOLUTION PATH FOR INTERPRETATION

**Parameter Estimates for Centered and Scaled Predictors**

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std Error</th>
<th>ChiSquare</th>
<th>Prob &gt; ChiSquare</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.655</td>
<td>0.0050049</td>
<td>17127.548</td>
<td>&lt;.0001*</td>
<td>0.6451906</td>
<td>0.6648094</td>
</tr>
<tr>
<td>Catalyst(1.2)</td>
<td>0.0975</td>
<td>0.0050049</td>
<td>379.50877</td>
<td>&lt;.0001*</td>
<td>0.0876906</td>
<td>0.1073094</td>
</tr>
<tr>
<td>Temperature(140.180)</td>
<td>0.05375</td>
<td>0.0050049</td>
<td>115.33723</td>
<td>&lt;.0001*</td>
<td>0.0439406</td>
<td>0.0635594</td>
</tr>
<tr>
<td>Catalyst*Temperature</td>
<td>0.06625</td>
<td>0.0050049</td>
<td>175.22027</td>
<td>&lt;.0001*</td>
<td>0.0564406</td>
<td>0.0760594</td>
</tr>
<tr>
<td>Temperature*Concentration</td>
<td>-0.055</td>
<td>0.0050049</td>
<td>120.76413</td>
<td>&lt;.0001*</td>
<td>-0.064809</td>
<td>-0.045191</td>
</tr>
<tr>
<td>Scale</td>
<td>0.0283119</td>
<td>0.002729</td>
<td>107.62733</td>
<td>&lt;.0001*</td>
<td>0.0229631</td>
<td>0.0336607</td>
</tr>
<tr>
<td>Concentration(3,6)</td>
<td>-0.03125</td>
<td>0.0050049</td>
<td>38.986355</td>
<td>&lt;.0001*</td>
<td>-0.041059</td>
<td>-0.021441</td>
</tr>
<tr>
<td>Catalyst*Concentration</td>
<td>0.01</td>
<td>0.0050049</td>
<td>3.9922027</td>
<td>0.0457*</td>
<td>0.0001906</td>
<td>0.0198094</td>
</tr>
<tr>
<td>Feed Rate(10,15)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0000</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
INTERPRETING THE SOLUTION PATH

• The parameter paths are selectable and are dynamically connected to the report.

• Move the black arrow to change the model being viewed. The entire report, including all graphics and tables, updates immediately.

• Blue lines are coefficients in the current model; black ones are zero-valued coefficients not in the current model.

• The parameter paths show strength and direction of the relationship with the response.

• The shape of the lines gives interesting information about the design. In this case, the lines are constant, which means the design is orthogonal.
• We see a range of models (Steps 5-9) within the green and yellow zones. These models have good support from the data.

• There is almost no difference between Step 6 (the best model) and Step 5, which differ by Catalyst*Concentration. Although it is marginally significant, we might consider dropping it from the model.

• Interactively changing the model in the zones in combination with the Profiler and Actual by Predicted plots does not show big changes.

• A combination of goodness of fit, sensible model parsimony, and subject matter knowledge should be used to determine the final model.
Non-Normal distributions are common, but are not part of the traditional DOE training.

They happen often when the response is strictly positive, a success/failure binary, a count.

Greater variation for larger values of the response is often best explained by non-normality.

The old-school approach would be to transform the response.

A modern, unified approach is to fit non-normal distributions and choose one based on the model selection criteria and your subject matter knowledge.

This is just like how we do variable selection!
NON-NORMAL DISTRIBUTIONS

- Cauchy – Outliers
- Binomial – Binary and nSuccess out of nTrials.
- Poisson – Count data
- Beta – Proportions (0,1)
- Gamma, Exponential - (0,∞)
- ZI – “Zero-Inflated”
- Beta, Binom, Neg. Poisson – “Overdispersed” count data.
• Reactor’s response is a proportion. Predictions outside (0,1) are meaningless.

• The Beta distribution is a possible alternative to the Normal distribution.

• The best Beta AICc is -100, vs. -115 for the Normal. The Normal Predictions stay in (0,1) range. I would stay with the Normal, but it is easy and worthwhile to take a look.
The Profiler is an extremely useful tool for extracting information about a model.

It shows traces (profiles) of the prediction formula with respect to each input variable, holding the other ones constant.
**THE PROFILER**

- The Profiler is where one:
  - Extracts predictions and prediction intervals from a model.
  - Optimizes a model, possibly with constraints.
  - Assess variable importance.
What are the most important variables in our model?

There are several related statistical tools for this:

- **Sums of Squares**: How much variation in the data is explained by a variable (or interaction, squared term)?

- **P-Value**: How likely is that you would see a larger coefficient than the one observed if the “true” one is zero?

- Neither of these tools directly tells us what are the most important variables in the model.
• Example: A regression coefficient can be highly significant with $p<.0001$ but still be very small in impact on the function that has been fit to the data (small coefficient, very small standard error).

• Another problem is that measures of variable importance tend to reflect the structure of the model and often don’t generalize to other models.

• A method like sums of squares works well for linear models, but is not intended for binary response models, PLS models, or Neural models.
SOBOL’S SENSITIVITY INDICES

- Sobol’s Sensitivity Indices are a general method for quantifying the amount of variability of a general function due to each of the inputs.

- Based on a decomposition of a function with regard to a probability density, $\mu(x_1, x_2, ..., x_k)$.

$$f(X) = f_0 + \sum_{i=1}^{k} f_i (x_i) + \sum_{i<j}^{k} f_{ij} (x_i, x_j) + \cdots + f_{12...k} (x_1, x_2, ..., x_k)$$

- The functions, $f_i$, $f_{ij}$, etc. are the marginal models and are orthogonal wrt probability measure $\mu(x_1, x_2, ..., x_k)$. 
SOBOL’S SENSITIVITY INDICES

\[ f(X) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_{i<j}^{k} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots k}(x_1, x_2, \ldots, x_k) \]

Where, for example:

\[ f_0 = \int f(x_1, \ldots, x_k) \, d\mu(x_1, x_2, \ldots, x_k) = E(f(x_1, \ldots, x_k)) \]

(overall average)

\[ f_1 = \int f(x_1, x_2, \ldots, x_k) \, d\mu(x_2, \ldots, x_k) - f_0 = E(f(x_1, \ldots, x_k)|x_1) - f_0 \]

(marginal \( x_1 \) main effect)

\[ f_{12} = \int f(x_1, x_2, \ldots, x_k) \, d\mu(x_3, \ldots, x_k) - f_0 - f_1(x_1) - f_2(x_2) \]

\[ = E(f(x_1, \ldots, x_k)|x_3, \ldots, x_k) - f_0 - f_1(x_1) - f_2(x_2) \]

(marginal \( x_1 \) \( x_2 \) interaction effect)
The idea is that the variability in the function can be uniquely decomposed into sums of squares attributable to each of these main effects and interaction terms. For example,

- \( SSQ_i = \int f_i^2(x_i) \, d\mu(x_i) = \text{Var}(E(f_i(x_i) \mid x_i)) \)
- \( SSQ_{total} = \int (f(x_1, \ldots, x_k) - f_0)^2 \, d\mu(x_1, \ldots, x_k) = \text{Var}(f(x_1, \ldots, x_k)) \)
- \( S_i = \frac{SS_i}{SS_{total}} = \frac{\text{Var}(E(f_i(x_i) \mid x_i))}{\text{Var}(f(x_1, x_2, \ldots, x_k))} \)

is the proportion of the variability due to \( x_i \) acting alone.

- We call this the main effect importance of \( x_i \).

- We can similarly define interaction effect importances of any order.
We measure the total impact of a variable by calculating the loss of variation that results from integrating it out:

\[ f_{-1} = \int f(x_1, ..., x_k) \, d\mu(x_1) - f_0 = E(f(x_1, ..., x_k) | f(x_2, ..., x_k)) \]

\[ SSQ_{-1} = \int f_{-1}^2(x_2, ..., x_k) \, d\mu(x_2, ..., x_k) \]
\[ = Var(E(f(x_1, ..., x_k) | f(x_2, ..., x_k))) \]

\[ S_{-1} = (SSQ_{total} - SSQ_{-1}) / SSQ_{total} \]
\[ = 1 - Var(E(f(x_1, ..., x_k) | f(x_2, ..., x_k)))/Var(f(x_1, x_2, ..., x_k)) \]

is the proportion of the variability lost due to integrating \( x_1 \) out.

- \( S_{-1} \) implicitly takes into consideration the main effect of \( x_1 \) and all of its higher order interactions!

- We call this the total effect importance of \( x_i \)
• One of the great things about these importances is that they make very few assumptions about function.

• The same technique can be applied to linear models, response surface models, logistic models, neural networks, PLS models, tree-based models, and model averaged models!

• Although there is quite a bit of math behind the scenes, the results are easy to use and interpret.
JMP uses Monte Carlo (until the standard error is 1% for all indices) to compute the integrals.

There are four options for the Monte Carlo distribution:

- Independent Uniform
  - Good for DOEs without constraints.
- Independent Resampled (from the data)
  - Fast for observational data, ignores multicollinearity.
- Dependent Resampled
  - Slower, but takes into account multicollinearity.
- Linearly Constrained Inputs
  - Uniform over linearly constrained region, only for DOEs with constraints (e.g., mixture designs), prevents extrapolation out of design region.
Nitrogen Oxides (NOx) are toxic greenhouse gases that are common by-products of burning organic compounds.

An experiment was done on an industrial burner to control the amount of NOx it created.

A 32 run I-Optimal RSM design was created with 7 continuous factors:
- Hydrogen Fraction in primary fuel
- Air/Fuel Ratio
- Lance Position X
- Lance Position Y
- Secondary Fuel Fraction
- Dispersant
- Ethanol Percentage in primary fuel
In many biological and chemical experiments, there is a smallest reading below which a reading is considered inaccurate. This is called a *lower limit of detection* (LOD) on the response.

A simple approach is to enter zeros for the readings at or below the LOD. This leads to flawed, biased results.

The better way to do the analysis is to use censoring.

A censored observation is one that we only observe to be within a certain (possibly infinite) range.
• There are three types of censoring: right, interval, and left censoring.

• Right censoring is very common in engineering reliability and in clinical studies where the response is the time to an event.

• For example, if a patient is in a 30-day study that evaluates a medicine that prevents migraines, and the study ends before the patient’s next migraine, then the recording would be a observation that is censored at 30 days. All we know is that the time until the next migraine was longer than 30 days, which should be reflected in a proper analysis.
• LOD data is left censored: If a measurement comes in at or below the LOD, all we know is that the actual value is somewhere between the lower detection limit and zero.

• Typically LOD data is strictly positive. This means that the data should be analyzed with a non-Gaussian distribution to avoid negative values predictions and variance heterogeneity.

• Analyzing LOD data in JMP is simple, you just have to have the response saved properly.
• To represent LOD data in JMP, you need *two response columns*: a low value and a high value.

• The two columns are the same for values above the LOD.

• Data below the LOD have a missing low value and a high value equal to the LOD.
• Rows 1, 2, and 5 are above the LOD, while rows 3 and 4 were at or below the LOD.

<table>
<thead>
<tr>
<th></th>
<th>Y Lo</th>
<th>Y Hi</th>
<th>Below LOD?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.5</td>
<td>10.5</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>7.1</td>
<td>7.1</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>•</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>•</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>11.2</td>
<td>11.2</td>
<td>No</td>
</tr>
</tbody>
</table>
• Researchers wanted to optimize determination of a pesticide (Metacrate) from water using Dichloromethane and Methanol as a dispersive and a solvent.

• They created a 32 run I Optimal design in JMP using Dichloromethane, Methanol, and Water Sample Volume as inputs.

• Four of the 32 observations were below the LOD of 1.0.
<table>
<thead>
<tr>
<th></th>
<th>Dichloromethane</th>
<th>Methanol</th>
<th>Sample Volume</th>
<th>Metacrate Original Reading</th>
<th>Metacrate (Lo)</th>
<th>Metacrate (Hi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>110</td>
<td>500</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>130</td>
<td>600</td>
<td>7</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>130</td>
<td>400</td>
<td>3</td>
<td>1.2</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>400</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>130</td>
<td>500</td>
<td>3</td>
<td>1.9</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>500</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>130</td>
<td>500</td>
<td>5</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>8</td>
<td>130</td>
<td>600</td>
<td>3</td>
<td>9.7</td>
<td>9.7</td>
<td>9.7</td>
</tr>
<tr>
<td>9</td>
<td>150</td>
<td>500</td>
<td>3</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>10</td>
<td>130</td>
<td>600</td>
<td>5</td>
<td>9.5</td>
<td>9.5</td>
<td>9.5</td>
</tr>
<tr>
<td>11</td>
<td>150</td>
<td>600</td>
<td>7</td>
<td>1.4</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>110</td>
<td>600</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>130</td>
<td>500</td>
<td>5</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>150</td>
<td>400</td>
<td>7</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>15</td>
<td>150</td>
<td>400</td>
<td>3</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>16</td>
<td>130</td>
<td>500</td>
<td>5</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>17</td>
<td>130</td>
<td>400</td>
<td>7</td>
<td>8.2</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>18</td>
<td>110</td>
<td>400</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>130</td>
<td>500</td>
<td>7</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>20</td>
<td>110</td>
<td>600</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>150</td>
<td>600</td>
<td>5</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>22</td>
<td>150</td>
<td>500</td>
<td>5</td>
<td>5.6</td>
<td>5.6</td>
<td>5.6</td>
</tr>
<tr>
<td>23</td>
<td>110</td>
<td>600</td>
<td>5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>110</td>
<td>400</td>
<td>7</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>