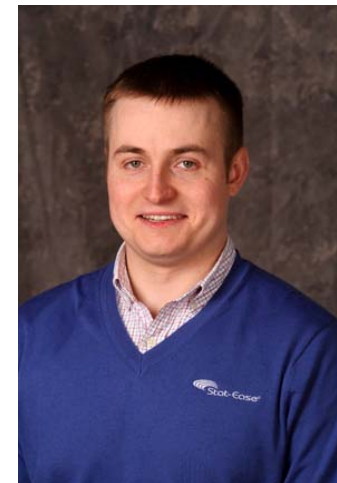


Taking Advantage of Automated Model Selection Tools for Response Surface Modeling

There are many attendees today! To avoid disrupting the Voice over Internet Protocol (VoIP) system, I will mute all. Please use the Questions feature on GotoWebinar. Feel free to email questions to stathelp@statease.com, which we will answer off-line.

-- *Martin*

*Presentation is posted at www.statease.com/webinar.html



Presented by Martin Bezener, PhD
Stat-Ease, Inc., Minneapolis, MN
martin@statease.com

Stat-Ease, Inc.



- Provider of DOE software, training & consulting
- Founded in 1982
- Headquartered in Minneapolis near the University of Minnesota
- Employs over a dozen professionals plus a worldwide network of resellers
- Mission is “Statistics Made Easy”
- Honored multiple times by American Society of Quality (ASQ) Chemical and Process Industries Division with Shewell Award recognizing excellence in presentation, scientific quality and applicability



Webinar Outline



- The DOE Process
- Model Selection: Why and How
- Software Demonstration
- Practical Tips and Tricks
- Resources and More Information

- **The DOE Process**
- Model Selection: Why and How
- Software Demonstration
- Practical Tips and Tricks
- Resources and More Information

The (Ideal) DOE Process



1. Formulate objective and goal
2. Determine factors of interest
3. Range finding (optional)
4. Screening (optional)
5. Characterization
6. Select model
7. Check model
8. Interpret/Optimize
9. Confirm optimization results

The (Ideal) DOE Process



1. Formulate objective and goal
2. Determine factors of interest
3. Range finding (optional)
4. Screening (optional)
5. Characterization
- 6. Select model** ← Focus of this webinar
7. Check model
8. Interpret/Optimize
9. Confirm optimization results

Model Selection: A Simple Example



- Suppose we run a central-composite design with the following factors and responses:
 - Factor **A: Temperature** (300 – 600 deg C)
 - Factor **B: Time** in Oven (20 – 40 minutes)
 - Response: Hardness

- A full quadratic model:

$$\text{Hardness} = I + A + B + AB + A^2 + B^2$$

is fit. There are 5 non-intercept terms in the model.

Model Selection: A Simple Example

(continued)

ANOVA for Response Surface Quadratic model					
Analysis of variance table [Partial sum of squares - Type III]					
Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	1275.01	5	255.00	63.44	< 0.0001
<i>A-Temperatu</i>	437.26	1	437.26	108.79	< 0.0001
<i>B-Time</i>	101.50	1	101.50	25.25	0.0015
<i>AB</i>	257.12	1	257.12	63.97	< 0.0001
<i>A²</i>	2.71	1	2.71	0.67	0.4387
<i>B²</i>	459.07	1	459.07	114.22	< 0.0001
Residual	28.14	7	4.02		
<i>Lack of Fit</i>	13.51	3	4.50	1.23	0.4077
<i>Pure Error</i>	14.62	4	3.66		
Cor Total	1303.15	12			

- The p-value for **A²**, the quadratic effect of **A**, is large (> 0.05), so there's little evidence of **A** affecting the response in a quadratic manner. We can remove this term and refit the model. **Easy Enough!**

Model Selection: A Simple Example

(continued)

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	1272.30	4	318.08	82.50	< 0.0001
<i>A-Temperatu</i>	437.26	1	437.26	113.41	< 0.0001
<i>B-Time</i>	101.50	1	101.50	26.33	0.0009
<i>AB</i>	257.12	1	257.12	66.69	< 0.0001
<i>B²</i>	476.42	1	476.42	123.57	< 0.0001
Residual	30.84	8	3.86		
<i>Lack of Fit</i>	16.22	4	4.06	1.11	0.4612
<i>Pure Error</i>	14.62	4	3.66		
Cor Total	1303.15	12			

- By taking out A^2 , we note the following. **Coincidence?**
 - Lack of fit p-value: 0.4077 \Rightarrow 0.4612
 - Adjusted R^2 : 0.9630 \Rightarrow 0.9645
 - Predicted R^2 : 0.9087 \Rightarrow 0.9249

Model Selection: A Complicated Example



- Suppose we set up a combined mixture-process design to optimize coffee flavor, with the following factors:
 - 3-Component Mixture:
 - Light Beans (0 – 100 wt%)
 - Medium Beans (0 – 100 wt%)
 - Dark Beans (0 – 100 wt%)
 - Process Factor: Amount of Coffee (2.5 to 4.0 ounces)
 - Categorical Factor: Grind Size (Fine, Medium, Coarse)
- A fully-crossed quadratic x quadratic model has **41** potential terms!
- If we had used a 5-component mixture, the full model would have contained around **105** terms. **Not easy** to select a model manually!

Webinar Outline



- The DOE Process
- **Model Selection: Why and How**
- Software Demonstration
- Practical Tips and Tricks
- Resources and More Information

- **Ultimate Goal in RSM:** Build a model which adequately describes the effect that the factors have on the response, then use that model for optimization.
- **Remember:** All models are wrong, some are useful!
- There may be factors that you varied in the experiment that don't actually affect the response in a meaningful way.
- In other cases, some interactions or higher-order effects (quadratic, cubic, etc.) may not be present.
- **Goal:** Build a model that contains the important terms.

Some Example Models



Situation: Predicting a person's weight without actually being able to weigh them.

- A model that's **too small**:
 - $\text{Weight} = f(\text{Age})$
- A model that's **small** and probably **reasonable**:
 - $\text{Weight} = f(\text{Age}, \text{Height}, \text{Daily Calories})$
- A model that's **larger** and probably **reasonable**:
 - $\text{Weight} = f(\text{Age}, \text{Height}, \text{Daily Calories}, \text{Daily Exercise}, \text{Waist Size})$
- A model that's **excessively large** and **not reasonable**:
 - $\text{Weight} = f(\text{Age}, \text{Height}, \text{Daily Calories}, \text{Daily Exercise}, \text{Waist Size}, \text{Eye Color}, \text{Shoe Size}, \text{Hair Color}, \text{Car Color}, \text{401k Balance}, \text{First Digit of Zip Code}, \text{Wears Glasses (Y/N)}, \text{Likes Rap Music (Y/N)}, \dots)$

Model Reduction: Why?



- If you have two reasonable models, one of which is **small** and the other which is **large**, the smaller model will offer several advantages:
 - Smaller models are “simpler”.
 - Smaller models are easier to interpret.
 - It’s easier to see how each factor affects the response.
- Large models are not necessarily bad, but a model that’s **excessively large** may encounter issues with:
 - Collinearity (redundancy)
 - Poorly estimated model coefficients
 - Generality (irreproducible results)

Idea Behind Automatic Model Selection



- As we saw earlier, there may simply be too many terms to deal with, which makes manual model selection difficult.
- You may also run into a situation where there are more model terms under consideration than runs in the experiment. **Cannot fit all the terms simultaneously!**
- **Basic Idea:** Let the computer pick a model for us.
- Let the computer try fitting several models, and then pick the “best” one.
- **Question:** How does automatic model selection actually work?

Let's Find Out!

How Automatic Model Selection Works

(Forward Selection)



- Suppose we have four terms the in **candidate set**: A, B, AB, B².
- **Forward selection**: Start with a “base” model. The program adds the ‘best’ term one at a time, until there are no more ‘good’ terms left to add.
 - The following models are fit:
 - I + A
 - I + B
 - I + AB
 - I + B²
 - The term that produces the best improvement (if any) is added to the model. Suppose it’s **B** in this illustration.

How Automatic Model Selection Works

(Forward Selection)



- The following models are now fit:
 - $I + B + A$
 - $I + B + AB$
 - $I + B + B^2$
- The term that produces the best improvement (if any) is added to the model.
- This process continues until the model cannot be improved by adding terms.
- Example Sequence:

$$I \Rightarrow I + B \Rightarrow I + B + A$$

How Automatic Model Selection Works

(Backward Selection)



- Suppose we have four terms the in **candidate set**: A, B, AB, B².
- **Backward selection**: The full model is fit. The 'worst' term is dropped one at a time, until only 'good' terms are left.
 - Fit the full model: I + A + B + AB + B².
 - The following models are fit:
 - I + ~~A~~ + B + AB + B²
 - I + A + ~~B~~ + AB + B²
 - I + A + B + ~~AB~~ + B²
 - I + A + B + AB + ~~B²~~
 - The term that produces the best improvement when dropped (if any) is chosen and dropped from the model.
 - Suppose in this case it's B², so the model is now I + A + B + AB.

How Automatic Model Selection Works

(Backward Selection)



- The following models are fit:
 - $I + \color{red}{\times} + B + AB$
 - $I + A + \color{red}{\times} + AB$
 - $I + A + B + \color{red}{\times}$
- The term that produces the best improvement when omitted (if any) is chosen and dropped from the model.
- This process continues until the model cannot be improved by dropping further terms.
- Example Sequence:

$$I + A + B + AB + B^2 \Rightarrow I + A + AB + B^2 \Rightarrow I + A + B^2 \Rightarrow I + A$$

(B Removed) (AB Removed) (B² Removed)

How Automatic Model Selection Works

(All Subset Selection)



- Suppose we have four terms the in **candidate set**: A , B , AB , B^2 .
- **All Subsets**: The program fits every possible model and picks the 'best' one.
 - The following models are fit:
 - I
 - $I + A$, $I + B$, $I + AB$, $I + B^2$
 - $I + A + B$, $I + A + AB$, $I + A + B^2$, $I + B + AB$, $I + B + B^2$, $I + AB + B^2$
 - $I + A + B + AB$, $I + A + B + B^2$, $I + A + AB + B^2$, $I + B + AB + B^2$
 - $I + A + B + AB + B^2$
 - The best of the 16 models is chosen.
 - Can be **very** time consuming if there are a lot of terms in the candidate set.

Model Selection Criteria



- The program cannot intuitively eyeball a “good” model, a “good” model term, the “worst” model term etc. – it needs a numeric score to compare models.
- There are many model scoring methods available. Most of them have the following form:
 - Score = goodness of fit (😊) + model size penalty (💣!)
- Why penalize the model size?
 - Adding model terms, even if unrelated to the response, will still improve the goodness of fit. However, some terms may not improve the goodness of fit enough to be worth including in the model.
- After scoring each model, the one with the best score is chosen.

Popular Criteria

- So we need to come up with a score. Or do we?
- There are many well-established model scoring methods out there. Three very popular methods for model selection include:
 - **BIC**: Bayesian Information Criterion (p = # of terms in model, n = # of runs)
 - $$\text{BIC} \approx \underbrace{n \times \ln(\text{ResidSS})}_{\text{Goodness of Fit}} + \underbrace{\log(n) \times p}_{\text{Size Penalty}}$$
 - **AIC**: Akaike Information Criterion (p = # of terms in model, n = # of runs)
 - $$\text{AIC} \approx \underbrace{n \times \ln(\text{ResidSS})}_{\text{Goodness of Fit}} + \underbrace{2 \times p}_{\text{Size Penalty}}$$
 - **AICc**: Corrected AIC for small samples
 - $\text{AICc} = \text{AIC} + \text{size correction}$
- Terms can also be included or excluded based on **p-values**.
- AICc and BIC: **lower** scores are better. We'll focus on AICc.

- Let's go back to the 2-factor central composite design with:
 - **A: Temperature** (300 – 600 deg C)
 - **B: Time** (20 – 40 mins)
- We'll consider the following model terms: A, B, AB, A², B².
- We'll try a few methods.
- Remember: When using AICc, BIC – lower is better!!

Advantages to Automatic Model Selection



- It's easy.
- It's fast.
- You can tailor the selection (via changing the criteria) to your particular problem.
- It produces a model that's usually at least somewhat reasonable.
- It doesn't require subject-matter knowledge (see next slide).

Disadvantages to Automatic Model Selection



- It doesn't use subject-matter knowledge.
- Different criteria and selection methods can produce different models. In some cases, the models may be very different. There's no one "correct" answer!
- Automatic selection, especially forward/backward selection, can be somewhat problematic when the design space is constrained or irregularly shaped (non-orthogonality of terms):
 - Order dependence
 - Correlated terms
 - Multiple testing/comparison issue (inflation of error rate)

Webinar Outline



- The DOE Process
- Model Selection: Why and How
- **Software Demonstration**
- Practical Tips and Tricks
- Resources and More Information

Design-Expert 10 Demonstration



- We already saw a demonstration of automatic selection on a relatively simple problem, where all methods produced the same model.
- We'll now briefly look at two (more complicated) examples to show how the Design-Expert® version 10 software can be used to do automatic model selection.
- We'll look at:
 - A constrained optimal RSM
 - A highly constrained optimal mixture design

Example 1

- A 3-factor constrained Optimal RSM design.

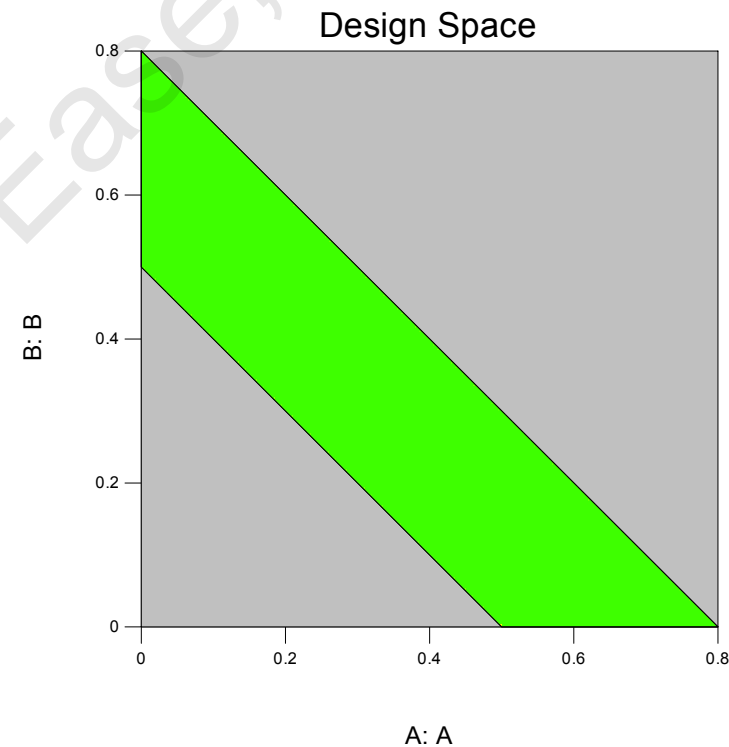
- $0 < \mathbf{A} < 0.8$

- $0 < \mathbf{B} < 0.8$

- $0 < \mathbf{C} < 2.0$

- Additional Constraint:

$$0.5 < \mathbf{A} + \mathbf{B} < 0.8$$



Example 1

(continued)

- Fit the full quadratic model.
- Remove everything but **C** ?

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	1137.88	9	126.43	19.44	< 0.0001
A-A	0.11	1	0.11	0.017	0.8995
B-B	1.98	1	1.98	0.30	0.5930
C-C	353.36	1	353.36	54.32	< 0.0001
AB	1.44	1	1.44	0.22	0.6479
AC	17.11	1	17.11	2.63	0.1359
BC	5.59	1	5.59	0.86	0.3758
A ²	0.10	1	0.10	0.016	0.9028
B ²	7.21	1	7.21	1.11	0.3172
C ²	4.74	1	4.74	0.73	0.4133
Residual	65.05	10	6.51		
Lack of Fit	29.97	5	5.99	0.85	0.5666
Pure Error	35.09	5	7.02		
Cor Total	1202.93	19			

Example 1

(continued)

- Let's use forward and backward selection with AICc, considering the following terms: **A, B, C, AB, AC, BC, A², B², C²**

- Forward:**

- $I + B + C + AC \Rightarrow I + A + B + C + AC$

Picked by DX10

Add **A** for hierarchy

- Backward:**

- $I + B + C + AB + AC + B^2 \Rightarrow I + A + B + C + AB + AC + B^2$

Picked by DX10

Add **A** for hierarchy

- Which model to use?!

Example 1

(continued)



- Compare the models:

	<u>Backward</u>	<u>Forward</u>
• LOF p-value:	0.6779	0.3599
• Predicted R2:	0.8661	0.7892
• AICc:	106.64	109.42

- Notice that the backward selection model is quite a bit better.
- Because this design space is constrained, the order in which the terms are added or dropped may cause a huge difference in your final model!

Example 1

(continued)

- The backward model:

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	1127.51	6	187.92	32.39	< 0.0001
A-A	3.604E-003	1	3.604E-003	6.213E-004	0.9805
B-B	9.08	1	9.08	1.56	0.2330
C-C	776.91	1	776.91	133.92	< 0.0001
AB	56.08	1	56.08	9.67	0.0083
AC	221.10	1	221.10	38.11	< 0.0001
B ²	58.95	1	58.95	10.16	0.0071
Residual	75.42	13	5.80		
Lack of Fit	40.33	8	5.04	0.72	0.6779
Pure Error	35.09	5	7.02		
Cor Total	1202.93	19			

- Note that almost all terms are significant! In the full quadratic model, only **C** was significant!! Order matters!

Example 1

(continued)



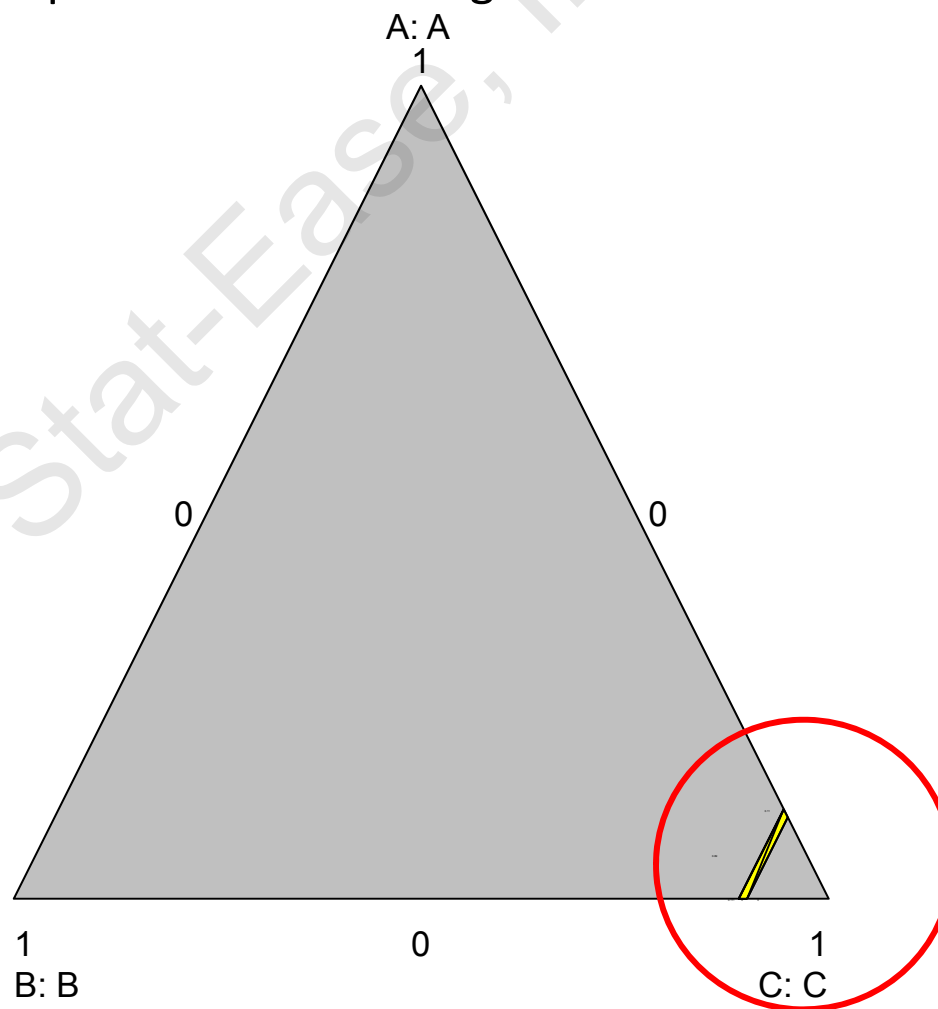
- Try maximizing the response using both models to check for any huge discrepancies.
- The maximum occurs at the following settings under each model:
 - **Forward:**
 - **A = 0.8, B = 0.0, C = 2.0** Predicted Response = 127.27
 - **Backward:**
 - **A = 0.8, B = 0.0, C = 2.0** Predicted Response = 129.93
- The maximum occurs at the same point in the design space. However, the predicted response using the backward model is probably more accurate.
- **Suggestion:** Perform 4-5 confirmation runs at the settings above.

Example 2

- A 3-component constrained Optimal Mixture design.

- $0 < \mathbf{A} < 0.11$
- $0 < \mathbf{B} < 0.11$
- $0.89 < \mathbf{C} < 0.90$
- Additional Constraint:

$$0.10 < \mathbf{A} + \mathbf{B} < 0.11$$



Example 2

(continued)

- Let's use forward and backward selection with AICc, considering the following terms: **A, B, C, AB, AC, BC, ABC, AB(A-B), AC(A-C), BC(B-C)**
- Forward:**

$$\mathbf{A + B + C + ABC + AB(A-B)}$$

Picked by DX10



Correct for hierarchy

$$\mathbf{I + A + B + C + AB + AC + BC + ABC + AB(A-B)}$$

- Backward:**

$$\mathbf{A + B + C + AC + BC + BC(B-C)}$$

Picked by DX10



Correct for hierarchy

$$\mathbf{A + B + C + AB + AC + BC + ABC + BC(B-C)}$$

Example 2

(continued)



- Compare the models:

	<u>Backward</u>	<u>Forward</u>
• LOF p-value:	0.5990	0.4064
• Predicted R2:	0.9965	0.9933
• AICc:	66.90	71.89

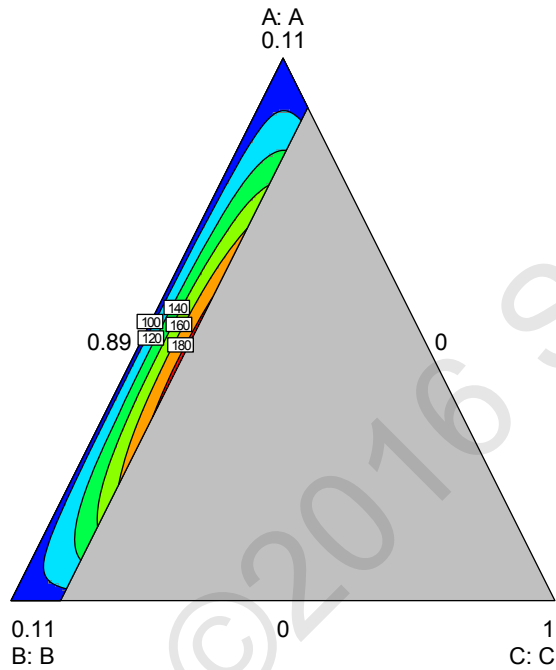
- Notice that the backward selection model is **a little** a bit better.

Example 2

(continued)

- Both models predict similarly:

Forward



Backward



Webinar Outline



- The DOE Process
- Model Selection: Why and How
- Software Demonstration
- **Practical Tips and Tricks**
- Resources and More Information

- **Forward** vs. **Backward** Selection
 - Why not try both if possible?
 - If both directions produce similar models, then you should have a good starting point.
 - If there are more potential model terms than runs in your data set, you must use forward selection.
 - There is some empirical evidence that backward selection does a bit better in cases where the design space is irregularly shaped or constrained.
 - Try **All Hierarchical Subset** selection if you don't have too many potential model terms.

- **AICc vs. BIC vs. p-values:**
 - AICc seems to be more popular in predictive modeling applications. BIC will often select too few terms.
 - The p-value method requires picking a threshold. The results may be very sensitive to whatever threshold you choose.
- If different methods produce different models, optimize using them and see if the results differ.
- Always use subject-matter knowledge! After using automatic model selection, ask yourself:
 - Do all of the terms in the model make sense?
 - Are there any excluded terms that should be in the model?
- Always perform confirmation runs to check model reproducibility.

Warning: R^2

- The coefficient of determination, R^2 , is sometimes used to do “model selection.”
- R^2 measures the % variability in the data that’s explained by the model.
- **BIG Problem:** R^2 can only get larger as you keep adding terms to your model, even if they are completely unrelated to the response. Larger models will be favored.
- **Do not** use R^2 to compare models of different sizes for model selection purposes.
- R^2 **can** be used to compare models of the same size.
- It’s safer to use the adjusted or predicted R^2 , which penalize excessively large models.

FYI: Other Criteria and Methods



- We've focused on forward and backward selection, using the AICc, BIC, and p-value criteria.
- Other model criteria & methods you may run into:
 - PRESS/Cross Validation
 - Predicted R^2
 - Mallows' C_p
 - Adjusted R^2
 - Regularized Regression (e.g. LASSO, Dantzig Selector, etc.)

Webinar Outline



- The DOE Process
- Model Selection: Why and How
- Software Demonstration
- Practical Tips and Tricks
- **Resources and More Information**

How to Get Help



- Search publications posted at www.statease.com.
- In Stat-Ease software press for Screen Tips, view reports in annotated mode, look for context-sensitive Help (right-click) or search the main Help system.
- Explore the Stat-Ease Experiment Design Forum <http://forum.statease.com> (read only).
- E-mail stathelp@statease.com for answers from Stat-Ease's staff of statistical consultants.
- Search for the [Stat-Ease YouTube Channel](#).

Support for DOE



- The triannual *Stat-Teaser* newsletter (if you don't opt out)
- Bi-Monthly *DOE FAQ Alert* e-mail (if you don't opt out)
 - Subscribe at: www.statease.com/doealertreg.html.
- StatsMadeEasy blog at www.statsmadeeasy.net.

Thank you for joining us today!

Best of luck in your future DOE work!

-Martin