Example 1
(continued)

- Compare the models:

	Backward	Forward
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:

ANOVA for Response Surface Reduced 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	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!

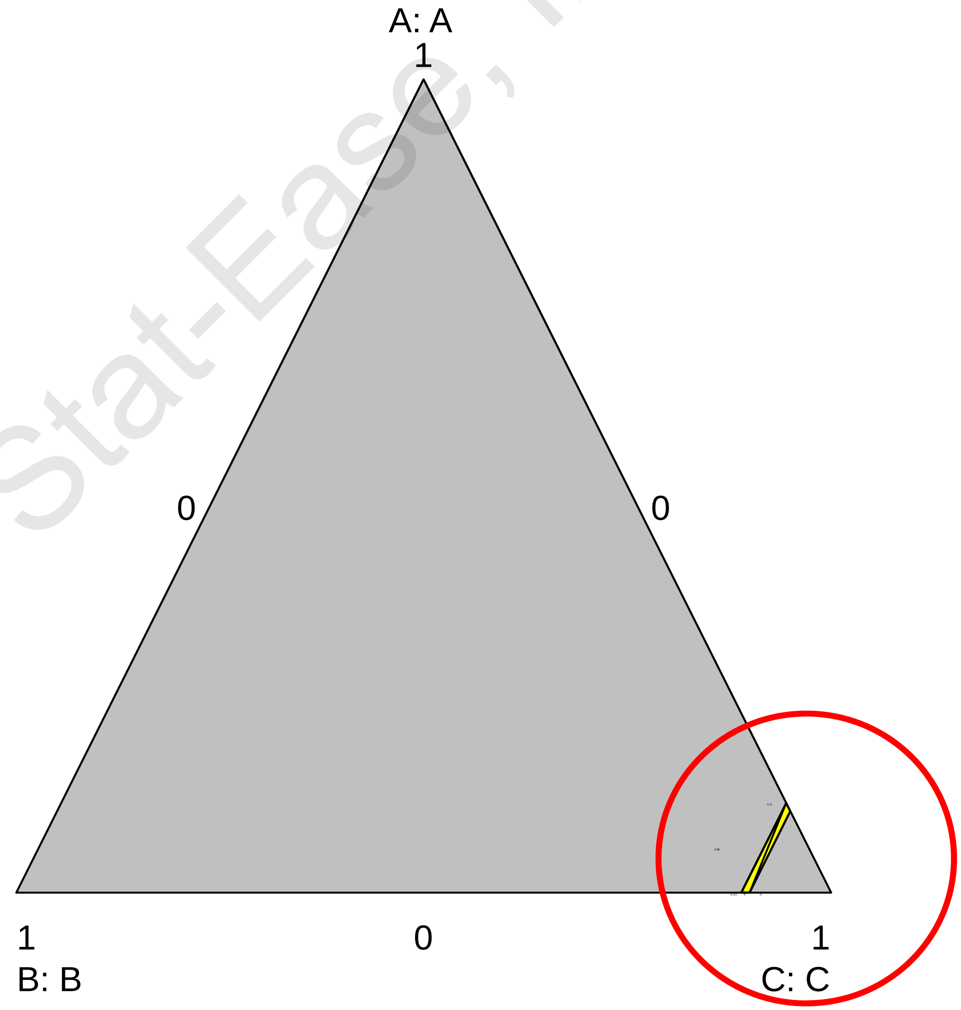
Example 1
(continued)

- 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:

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



Design Space

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:**

    A + B + C + ABC + AB(A-B)
    Picked by DX10

    Correct for hierarchy

    I + A + B + C + AB + AC + BC + ABC + AB(A-B)
  - **Backward:**

    A + B + C + AC + BC + BC(B-C)
    Picked by DX10

    Correct for hierarchy

    A + B + C + AB + AC + BC + ABC + BC(B-C)

Example 2
(continued)

- Compare the models:

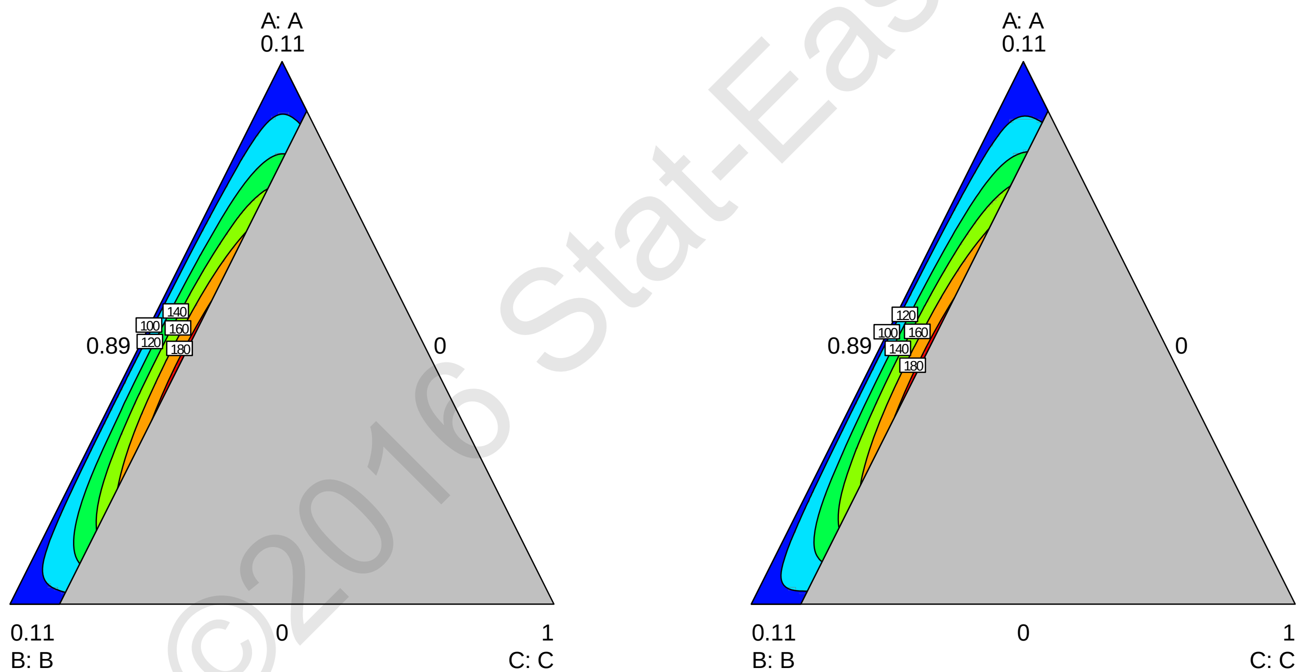
	Backward	Forward
LOF p-value:	**0.5990**	0.4064
Predicted R2:	**0.9965**	0.9933
AICc:	**66.90**	71.89

- 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

Practical Tips and Tricks 1

- **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.

Practical Tips and Tricks 2

- **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
  - Mallow's 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