Section 10 – Advanced Analysis Features

This section discusses a few selected topics on advanced tools for data analysis available in Design-Expert® software:

- Algorithmic model reduction backward, forward or stepwise
- Responses generated by an equation, for example: cost
- Propagation of error for robust response results

We encourage you to explore various features encountered as you work through this section. You're an experimenter, so it's good to be adventurous!

You may find more information in the Statistical Details: Analysis section of this manual. Also, check out the on-line Help system in the Design-Expert program. If you still need help, then give Stat-Ease a call. Our telephone number is in the Introduction.

Algorithmic Model Reduction

You can eliminate non-significant terms in one of two ways:

- Manually, by editing the list of candidate terms
- Automatically, by algorithmic techniques.

Design-Expert offers three automatic reduction algorithms:

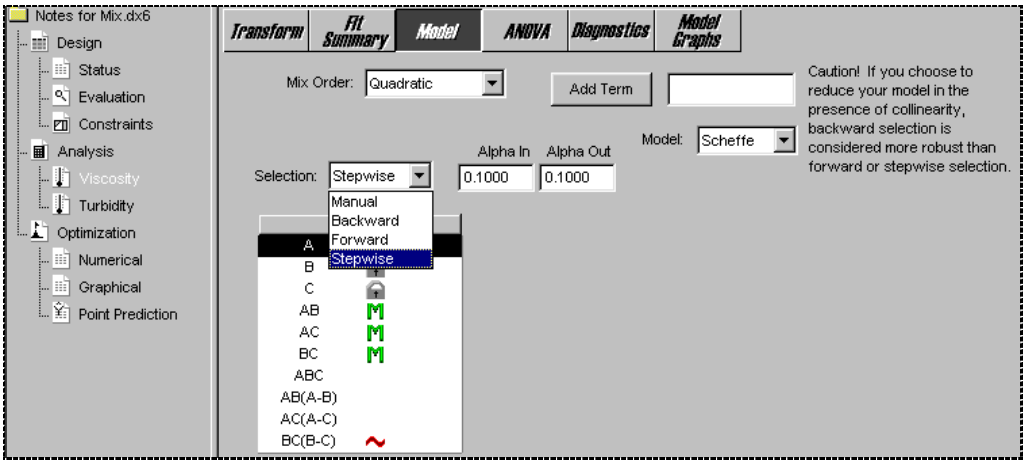
1. Backward elimination: Eliminate a factor at each step.
2. Forward selection: Add a factor at each step.
3. Step-Wise: Add, eliminate or exchange a factor at each step.

Since the step-wise method incorporates features of the other two approaches, it will be the main topic for discussion.

Step-Wise Regression

The use of model reduction will best be demonstrated by looking at a mixture design, which presents more of a challenge due to the inherent co-linearity between components. The case data from the mixture tutorials presented earlier in this manual will provide a good starting point for trying out the reduction tools. From the **File** menu, click **Open Design** and then double click on **Mix.dx6**.

After you have the file loaded, click on the **Viscosity** response node. From the analysis tool bar select ***Model***. For **Selection**, pick **Stepwise**. Leave the **Alpha In** and **Alpha Out** at their default values. Read on for more explanation.



Model Selection Screen: Stepwise Selection

You can change the model by a right mouse click on candidate terms. Choose to Model, Exclude or Force any individual term. Give this a try if you like. As indicated by the lock symbol, Design-Expert requires that the linear terms always be included in Scheffe mixture models.

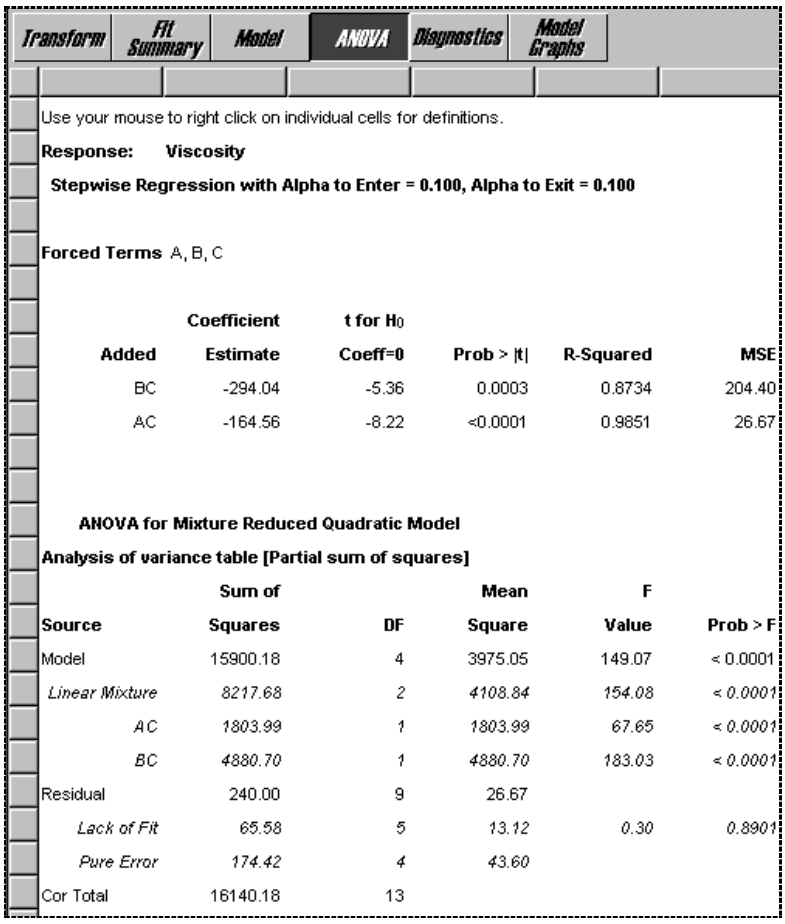
Also, note that you can add terms. The program provides an entry field for this purpose in the Model screen. It accepts terms up to 6^{th} order in total (for example: A^6 or A^3B^3 or ABC^4) for up to a certain number of factors or components. Check Help for details.

Now click on the ***ANOVA*** icon to do the step-wise regression. The algorithm adds the single new factor that exhibits the highest correlation with the response. After the first step, factors are added, eliminated, exchanged or the procedure stopped. When two or more terms are in the model, then Design-Expert will:

1. Add to the model the term having probability values less than the specified Alpha In requirement.
2. Remove all terms (one at a time) having probability values greater than the specified Alpha Out.
3. Stop when there is no further improvement and the Alpha In and Alpha Out values are satisfied.

The levels for Alpha In and Alpha Out are set higher than the final level of significance desired to ensure that more terms will be brought into the model.

After completing all these steps, Design-Expert displays an accounting of the added, along with key statistics.



Transform | Fit Summary | Model | ANOVA | Diagnostics | Model Graphs

Use your mouse to right click on individual cells for definitions.

Response: Viscosity

Stepwise Regression with Alpha to Enter = 0.100, Alpha to Exit = 0.100

Forced Terms A, B, C

Added	Coefficient Estimate	t for H_0 Coeff=0	Prob > |t|	R-Squared	MSE
BC	-294.04	-5.36	0.0003	0.8734	204.40
AC	-164.56	-8.22	<0.0001	0.9851	26.67

ANOVA for Mixture Reduced Quadratic Model

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	15900.18	4	3975.05	149.07	< 0.0001
Linear Mixture	*8217.68*	*2*	*4108.84*	*154.08*	*< 0.0001*
AC	*1803.99*	*1*	*1803.99*	*67.65*	*< 0.0001*
BC	*4880.70*	*1*	*4880.70*	*183.03*	*< 0.0001*
Residual	240.00	9	26.67		
Lack of Fit	*65.58*	*5*	*13.12*	*0.30*	*0.8901*
Pure Error	*174.42*	*4*	*43.60*		
Cor Total	16140.18	13			

Output from Stepwise Regression

You should note that the ANOVA results are very similar to those from the Mixture Tutorials section. Only the AB term is left out after the stepwise regression, providing a marginal improvement to the model.

Backward Elimination

Design-Expert does backward elimination as follows:

1. Begin with the full model.
2. Remove from the model the term with the highest partial probability value.
3. Stop when the probability value of the next term out satisfies the specified "Alpha Out" criterion.

The regression steps differ from stepwise but the results for this example remain the same. If you like, go ahead and try the backward approach now. In general, for well-designed experiments with minimal co-linearity, you will see no difference in results. However, the backward method may be the most robust choice since all model terms will be given a chance of inclusion in the model. Better yet, do a manual reduction guided by your knowledge of the product or process.

Forward Selection

Finally, if you like, try the forward approach. The results for this example will be identical to the step-wise and backward results.

The forward selection algorithm will follow these steps:

1. Begin with a simple regression model using the single factor having the highest correlation with the response.
2. Add to the model the factor with the lowest partial probability value.
3. Stop when the probability value of the next factor in does not meet the specified "Alpha In" criterion.

This algorithm may not be as robust as the others, because some terms may never get the chance to be included in the model. This will only cause concern if the data exhibits a high degree of collinearity.

Adding Responses via Equation

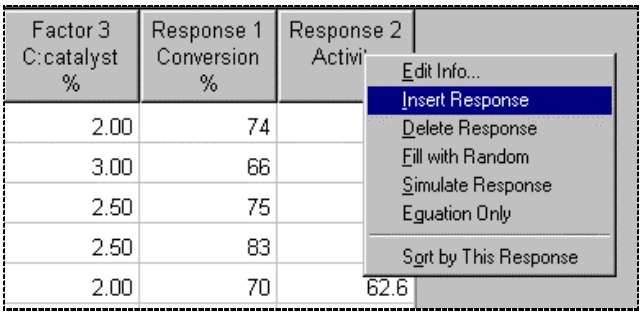
Design-Expert allows you to construct responses from your own models, computed exactly via equation-only or with error via simulation. The equation-only option will be especially useful for consideration of cost relationships. The simulation option allows you to test your skills at analysis. Stat-Ease uses this feature for its workshops. In either case you will be limited to standard polynomial forms. These can be modified by various transformations.

This section leads you through an example based upon the Response Surface Tutorials data set. First you will set up an equation that computes the cost of the product in terms of the actual factors in the design:

$$\text{Cost} = 5 + 1.0\text{Time} + 0.4\text{Temperature} + 60\text{Catalyst}$$

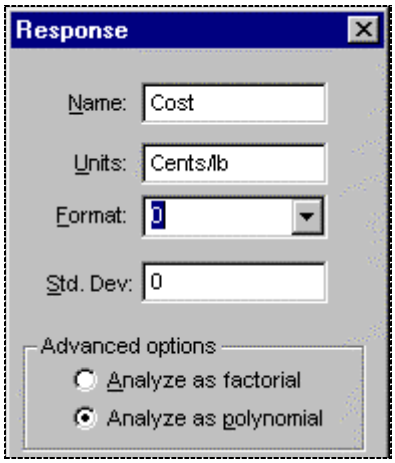
with cost in cents/pound, time in minutes, temperature in degrees Celsius and catalyst in units of percent.

Use **File**, **Open Design** to load **rsm-a.dx6**. Insert an empty response column by right clicking on the **Activity** response heading, and selecting **Insert Response**.



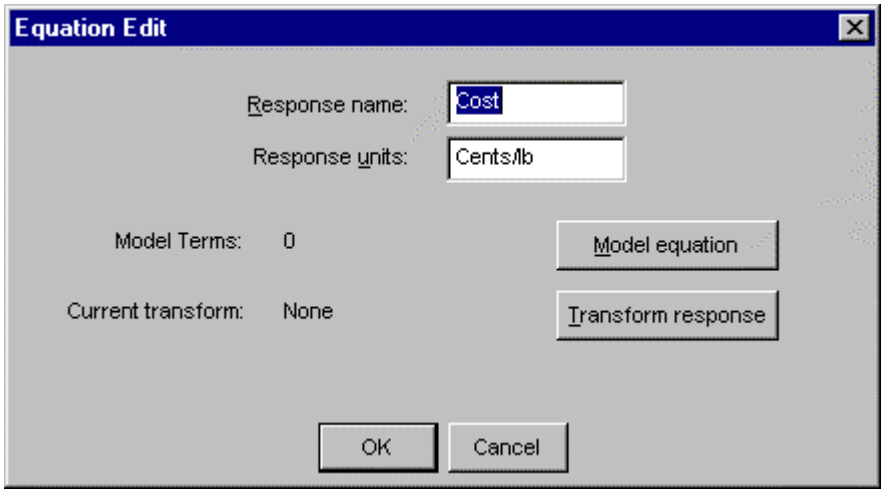
Inserting a Response Column

Right click on the new **Response** column header and choose **Edit Info**. Then enter information as shown below.



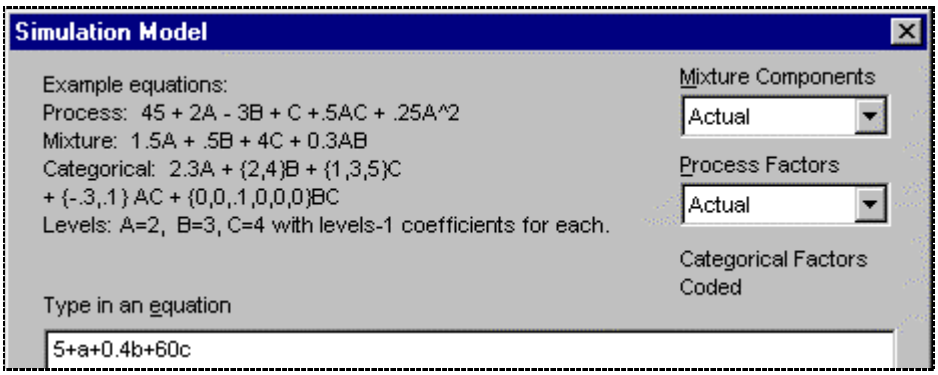
Edit Information Dialog Box for New Response

Leave the standard deviation (Std. Dev) at its default value. Also, don't change any of the listed options for "POE" (propagation of error), which will be discussed later. Press **OK**. Then right click again on the new response heading. This time select **Equation Only**.



Dialog Box for Entering an Equation

Now click on the **Model equation** button to bring up a form to enter/edit the model coefficients that describe the equation. Leave the defaults for variable coding as **Actual**. Now you are ready to enter the coefficients for the model, as shown earlier, but with lettered factors. Your screen should match that shown below.



Actual Cost Equation: Completed Entry

Accept the edited equation and subsequent screens by pressing **OK** twice. Design-Expert then inserts the calculated cost data into the response column.

Factor 1 A:time min.	Factor 2 B:temperature deg C	Factor 3 C:catalyst %	Response 1 Conversion %	Response 2 Activity	Response 3 Cost Cents/lb
40.00	80.00	2.00	74	53.2	197
40.00	90.00	3.00	66	59.8	261
45.00	85.00	2.50	75	60.4	234
45.00	85.00	2.50	83	60.6	234
50.00	90.00	2.00	70	62.6	211
45.00	85.00	2.50	76	59.1	234

Results of Cost Equation

You now can include the cost in your multiple response optimization. This allows you to find conditions to meet all specifications while minimizing the cost.

Propagation of Error

The propagation of error (POE) method finds settings that minimize variation in the response. It makes your process or product more robust to variations in input factors. In essence, the POE method involves application of partial derivatives to locate flat areas on the response surface, preferably high plateaus. For details on the mathematics, see "Robust Design - Reducing Transmitted Variation," a paper by Whitcomb and Anderson presented at the 50th Annual Quality Congress in 1996. (Contact Stat-Ease for a reprint.)

To use POE, you must first construct response surface models. Then you enter estimates of standard deviation for each input factor. With this information, Design-Expert software constructs a response surface map of transmitted variation. Finally, you can use multiple response optimization to find factor settings that get your response on target with minimal variation.

Refer back to the tutorials on response surface methods (RSM) and/or mixture design for detailed examples of how to apply POE. In the RSM case study, Design-Expert fitted the following predictive model to the conversion response:

$$\hat{Y} = 81.60 + 1.03 * A + 4.04 * B + 6.20 * C + 2.12 * AB + 11.38 * AC - 3.88 * BC - 1.90 * A^2 + 2.88 * B^2 - 5.25 * C^2$$

To make the POE plot, it then took the partial derivative to produce an equation that relates conversion variance to input factor variances.

$$\sigma_Y^2 = \left(1.03 + 2.12 * B + 11.38 * C + 2(-1.90) * A\right)^2 \sigma_A^2 + \left(4.04 + 2.12 * A - 3.88 * C + 2(2.88) * B\right)^2 \sigma_B^2 + \left(6.20 + 11.38 * A - 3.88 * B + 2(-5.25) * C\right)^2 \sigma_C^2 + \sigma_{resid}^2$$

The last term in the equation is the residual variance, which is estimated by the standard deviation of the response.

Note that for linear responses like activity, Design-Expert will not allow you to select Propagation of Error, because the calculations yield a constant value. However, linear responses like activity can be useful as adjustment factors to get your response back on target after you reduce transmitted variation. To learn more on how this works, attend the Stat-Ease workshop called Robust Design, DOE Tools for Reducing Variation. Call for details and schedule.

This completes the guided tour of advanced features. But don't stop exploring. You may find other handy features not deemed important enough to be included in this manual. You will find complete documentation at your fingertips via the Help system. You can then wander through the labyrinth of hypertext links to learn everything you want to know about Design-Expert but were afraid to ask. Enjoy!