Section 4 – General Factorial Tutorials

General Factorial Part One: Categorical

Design-Ease® software version 6 offers a "General Factorial" option on the "Factorial" tab. If you completed the One Factor Tutorial (recommended), you've seen how this option handles one multilevel, categorical factor. In this tutorial you will learn how to set up a design on multiple categorical factors.

Introduction

The following study, taken from Montgomery's *Design and Analysis of Experiments,* seeks increased life in a battery that will be subjected to extremes in ambient temperature. Three materials will be evaluated at three levels of temperature. Each experimental combination will be replicated four times in a completely randomized design. The responses from the resulting 36 runs can be seen below.

Material Type	Temperature (deg F)					
	15		70		125	
A1	130	155	34	40	20	70
	74	180	80	75	82	58
A2	150	188	136	122	25	70
	159	126	106	115	58	45
A3	138	110	174	120	96	104
	168	160	150	139	82	60

General Factorial on Battery (response is life in hours)

The following questions must be answered:

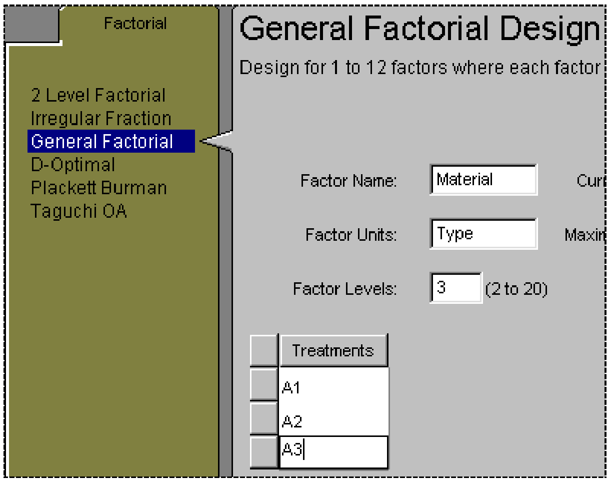
- How does material type and temperature affect battery life?
- Which material, if any, will give uniformly long life regardless of temperature?

The second question, if it can be answered in the affirmative, leads to the big payoff: a battery that will be robust to temperature variation in the field. This case study provides a good example of the application of statistical DOE for robust product design.

Design the Experiment

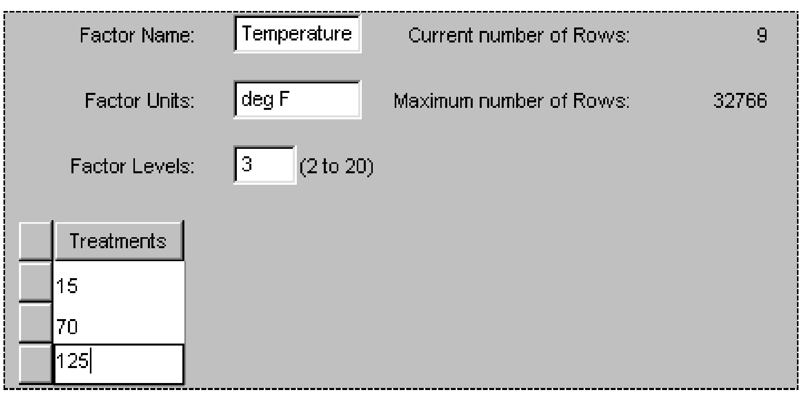
To build the design, click the blank-sheet icon 🗋 on the toolbar (or choose File, New Design). Then from the default **Factorial** tab, click **General Factorial**. Enter "**2**" as the number of factors. Click on the **Continue** button.

Enter "**Material**" for the factor name, "**Type**" for units and "**3**" for the number of levels. Change the treatment names to "**A1**", "**A2**" and "**A3**".



Entry Screen for Material Treatments

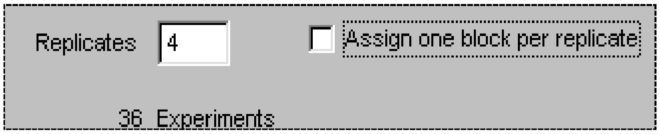
Press **Continue** to move on to the next factor. Enter: "**Temperature**" for the factor name, "**deg F**" for units, "**3**" for the number of levels, and "**15**, **70** and **125**" for the levels. Your screen should now look like that shown below.



Entry Screen for Temperature

Notice that the current number of rows (9 = 3 x 3) is far less than the maximum. However, you may hit this limit if you include too many factors at too many levels.

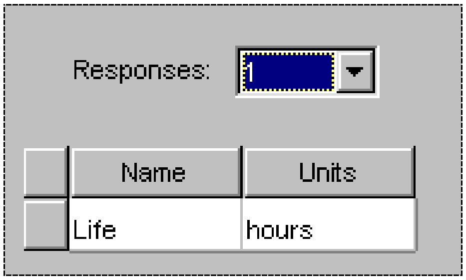
Click on the **Continue** button to complete the factor-entry stage of the design build. Enter "**4**" for replicates. (Recall from the table of data that four batteries were tested at every combination of material type and temperature.) Leave the blocks option alone, because these experiments are completely randomized.



Entering Number of Replicates

Click on the **Continue** button to move on to the entry screen for responses.

Leave the default responses at 1. Enter "**Life**" and "**hours**" for name and units.

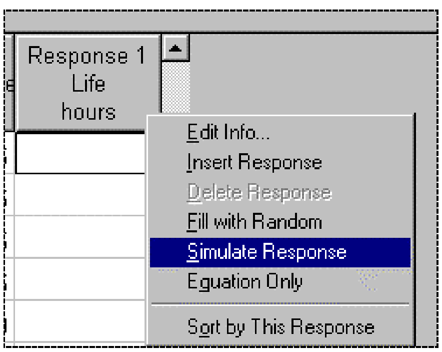


Response Entry Screen

Click on the **Continue** button. Design-Ease now displays the 36 runs (in random order) from the 3x3 factorial design with four replicates.

Analyze the Results

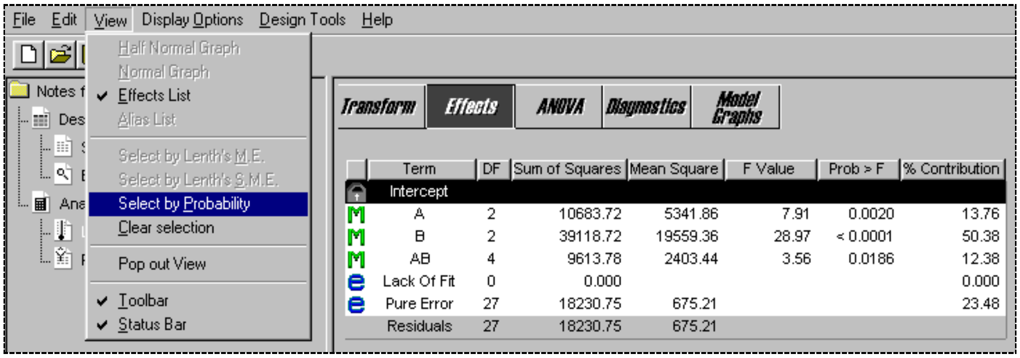
To save time, simulate the experimental results by right-clicking the response header and selecting **Simulate Response**.



Choosing a Simulation

Open the file named **Battery.sim**. You should now see data from the experiment. Go to **File** and **Save As**. Change the file name to **Battery.de6** and **Save**. To analyze the data click on the analysis node labeled "**Life**," which can be found in the tree structure along the left of the main window.

Click on the ***Effects*** button displayed in the toolbar at the top of the main window. From the main menu choose **View**, **Select by Probability**.



Effects View (After Selecting by Probability)

Notice that the significant factor effects have been chosen for the model (designated "M"). The term(s) labeled "e" will be used for error. (Note: The statistical criteria for selection can be configured by going to the default screen under Edit, Preferences.)

Click on the ***ANOVA*** button.

ANOVA for Selected Factorial Model

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Model	59416.22	8	7427.03	11.00	< 0.0001	significant
A	*10683.72*	*2*	*5341.86*	*7.91*	*0.0020*	
B	*39118.72*	*2*	*19559.36*	*28.97*	*< 0.0001*	
AB	*9613.78*	*4*	*2403.44*	*3.56*	*0.0186*	
Residual	18230.75	27	675.21			
Cor Total	77646.97	35				

The Model F-value of 11.00 implies the model is significant. There is only
a 0.01% chance that a "Model F-Value" this large could occur due to noise.

Values of "Prob > F" less than 0.0500 indicate model terms are significant.
In this case A, B, AB are significant model terms.
Values greater than 0.1000 indicate the model terms are not significant.
If there are many insignificant model terms (not counting those required to support hierarchy),
model reduction may improve your model.

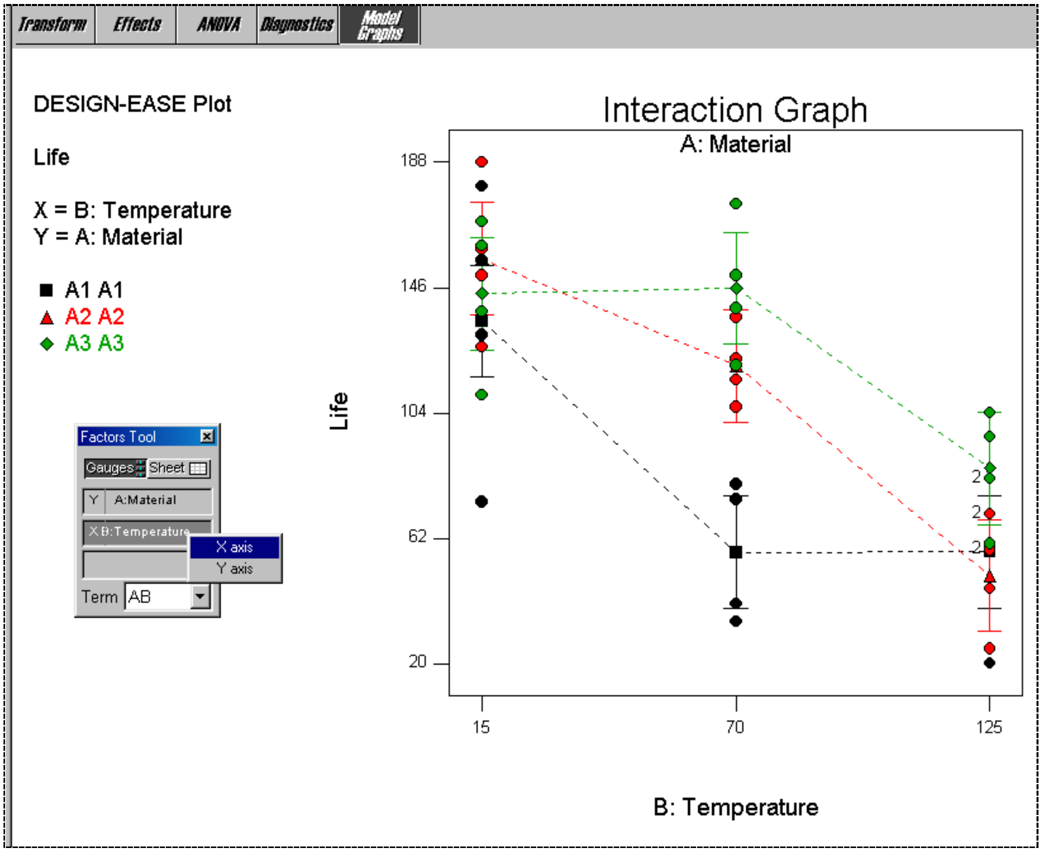
Std. Dev.	25.98	R-Squared	0.7652
Mean	105.53	Adj R-Square	0.6956
C.V.	24.62	Pred R-Squar	0.5826
PRESS	32410.22	Adeq Precisic	8.178

Annotated ANOVA Report

As you can conclude for yourself by reading the comments, the results look good. (If your screen does not display the comments, select View, Annotated ANOVA.)

Click on the ***Diagnostics*** button and examine the diagnostic graphs. These look OK.

Click on the ***Model Graphs*** button. Right click over the **Temperature** factor on the floating graph tool and change it to the **X-Axis**.

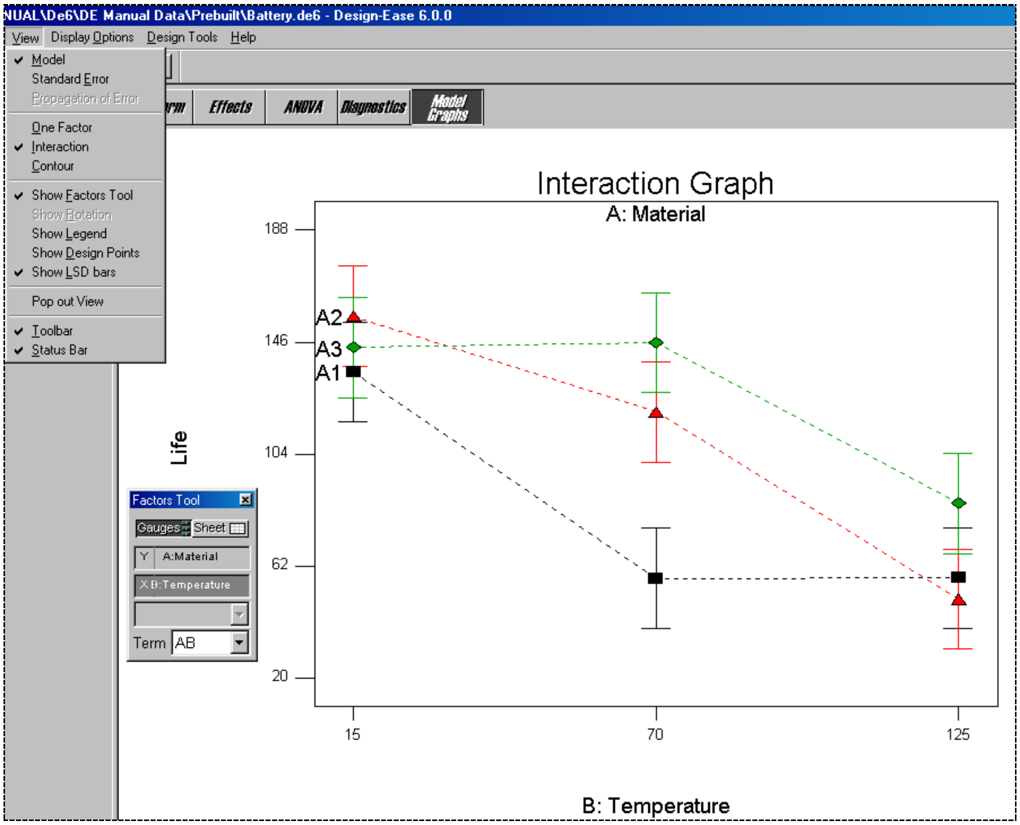


Effect Graph with Temperature as X-Axis (Point Clicked for ID Purposes)

You now see a graph with the continuous factor temperature on the x-axis, with discrete lines for the three categorical materials. Click on any round point for identification of an actual result, shown to the left of the plot. In the figure above, we highlighted the highest point: temperature of 15 with material A2, as noted to the left of the plot. (Note: due to randomization, your run number may differ for the chosen point.) You can also click on one of the other symbols (square, triangle or diamond) to display the predicted outcome and least significant difference (LSD). Try this!

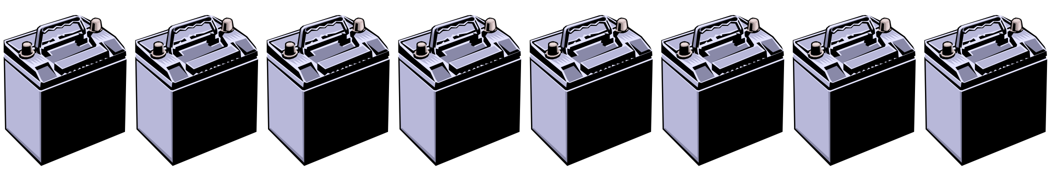
To produce a cleaner looking plot, go to **View** and deselect **Show Legend**. Go back to **View** and deselect **Show Design Points**. Your screen should now look like that shown below.

From this graph you can clearly see that all three materials work very well at the low temperature (15 degrees). Based on the overlapping LSD bars, it would be fair to say that no material stands out at this end of the scale. However, the A1 material clearly falls off at the 70 degree temperature, which would be encountered most often, so it must be rejected. None of the materials do very well at the highest temperature (125 degrees), but the upper end of the LSD bar for A2 barely overlaps the bottom end of the LSD bar for A3.



Cleaner Effect Graph with No Legend or Points

Therefore, in respect to temperature sensitivity, the experimenters concluded that material A3 is the most robust material for making batteries.



Split Plot (Advanced Topic)

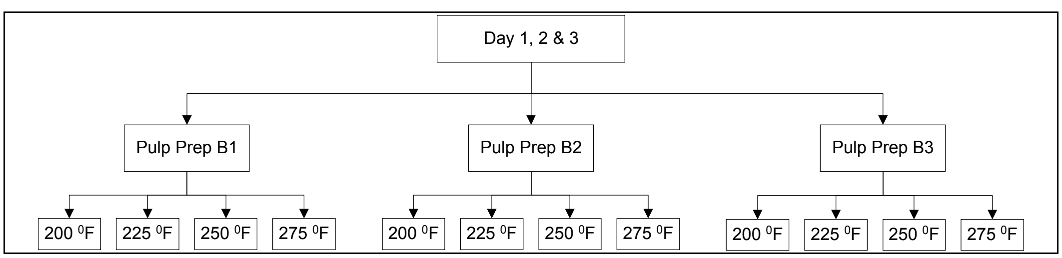
In some randomized block designs you must restrict the randomization. Otherwise it wouldn't be practical to perform the experiments. For example, temperature cannot be easily changed in many applications. The solution may be a "split-plot" design, which originated in the field of agriculture. Experimenters divided large areas of land, called "whole plots," into smaller "subplots" that could be treated separately. For example, they planted several crop varieties - one each per whole plot, and then applied different fertilizers on each subplot.

The analysis of a split plot design is tricky, even for statisticians. It can be done with Design-Ease by properly designating effects in specific ways for subsequent analysis of variance. Proceed if you dare!

To illustrate how Design-Ease software can be manipulated to do a split-plot, let's do an example from Montgomery's *Design and Analysis of Experiments*. A manufacturer wants to determine the effect of three methods of pulp preparation and four levels of cooking temperature on the tensile strength of paper. These twelve combinations can be done in one day.

To increase statistical power, the manufacturer decides to do three replicates of this general factorial design over three days, which produces a total of 36 runs. However, it will not be feasible to perform the experiment in a random fashion. Instead, on each of the three days, the experimenter produces a batch of pulp and divides it into four samples to be cooked at four different temperatures. This process is repeated for pulp prepared by the two other methods.

A flowchart of the experiment is shown below.



Flowchart of Split Plot Experiment on Papermaking Process

In statistical terms, the split plot experiment can be structured as:

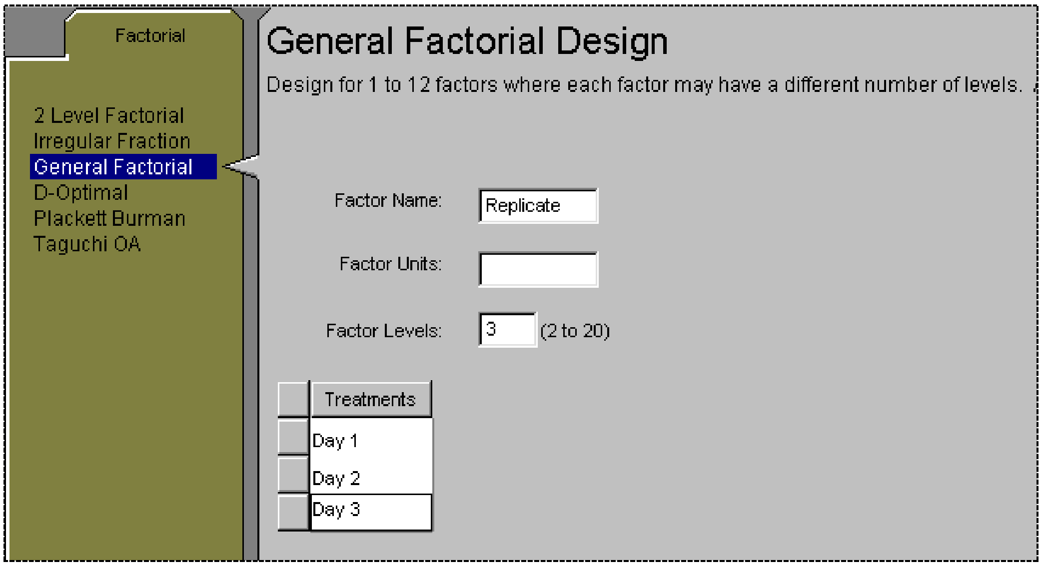
- Whole plots for the three batches of pulp
- Sub-plots for the four samples cooked at four different temperatures

You will set up this design as a general factorial. It will then need to be modified somewhat so it comes out in the right order. Then you will perform an analysis, which really gets tricky due to the restrictions in randomization.

Design the Experiment

Let's build the design. This is done via the usual procedure for general factorials, but with days explicitly accounted for as a factor. Click the blank-sheet icon on the toolbar (or choose File, New Design). Then from the default **Factorial** tab, click on **General Factorial**. Enter **3** as the number of factors. Click on the **Continue** button.

Next, enter **Replicate** for the factor name, **3** for the number of levels, **Day 1**, **Day 2** and **Day 3** for the treatment names. Your screen should now look like the illustration.



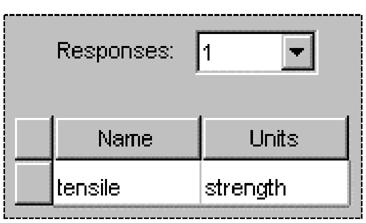
Entering Names of First Factor

Click on the **Continue** button to move on to the next factor. Then enter **Pulp Prep** for the factor name, and **3** for the number of levels. Change the treatment names to "**B1**", "**B2**" and "**B3**". Click on **Continue**.

Enter **Temp** for the final factor name, **deg F** for units, **4** for the number of levels and **200, 225, 250** and **275** for the levels. Click on the **Continue** button.

Leave "**1**" for Replicates and click on the **Continue** button.

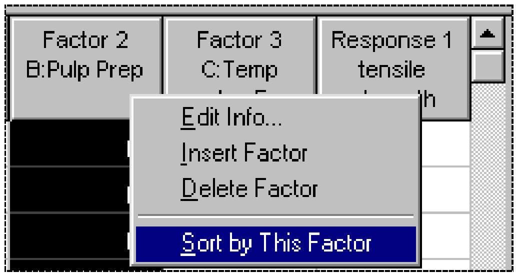
There is one response. Enter the name as "**tensile**" and the units as "**strength**" as shown below.



Entering Response Name

Click on the **Continue** button.

Design-Ease now displays the design in completely random order. Obviously this cannot be performed as listed - day 1 must precede day 2, all temperatures must be run on one method of pulp preparation before starting on the next, etc. To modify the order, first right click on the **Pulp Prep** (Factor 2) column heading and choose **Sort by This Factor**.



Sorting by Pulp Prep Method

Next, right click on the **Replicate** (Factor 1) column and select **Sort by This Factor**. Now you see subplot treatment (C - cooking temperature) randomized within the whole plot treatment (B - pulp preparation). To retain this order, right click on the **Run** column header and select **Re-order as currently displayed**.

Analyze the Results

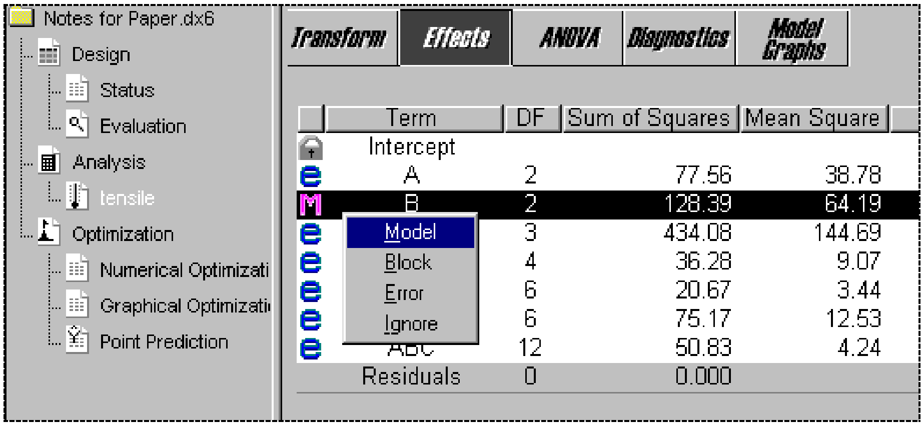
Read in the response data via **File**, **Open Design** from the main menu. Select the file named **Paper.de6**.

The analysis of a split plot design must be done somewhat manually. You need to create ANOVA's for the whole plot treatment (pulp prep), the subplot treatment (temperature) and the interaction between whole plot and subplot individually to get the correct statistical tests for each. Then you fit the full model in order to get meaningful diagnostics and model graphs (*but you ignore the full model ANOVA!*)

To analyze the whole plot treatment (B - pulp preparation) click on the analysis node labeled **tensile** which can be found in the tree structure along the left of the main window. Then click on the ***Effects*** button displayed in the toolbar at the top of the main window. There are four states an effect can have:

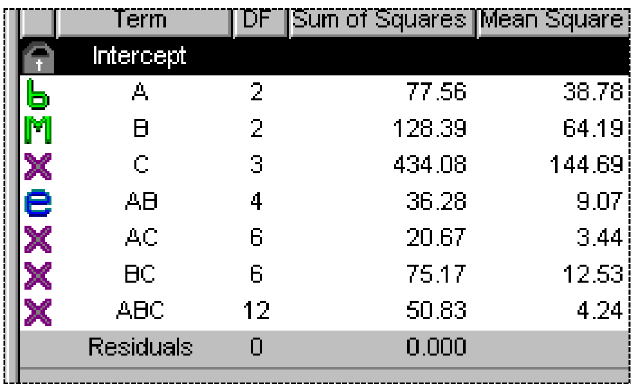
- Model ("M")
- Block ("b")
- Error ("e")
- Ignore ("X")

The whole plot treatment (B) should be tested against the Replicate by Pulp Preparation interaction, i.e., the AB interaction. In the effects list right click on factor **B** and choose **Model** by pressing **M**, or clicking on the word "Model." (Another way to designate a term for model is to double-click it. Try this on B if you like. It toggles back to error. Double click the term again to put it in the model.)



Right-click Menu for Designating Effects

We need to include a block correction in our model, so right click on factor **A** and choose **Block** (mark with "b"). Right click on **C** and **Ignore** it (mark with "X"). Do the same for **AC, BC and ABC** (suggestion: try clicking on AC and shift-clicking on ABC to highlight all three effects, and then do one right-click to designate this block of effects to be ignored). Leave **AB** at its default designation as **Error** ("e"). Your screen should now match the illustration below.



	Term	DF	Sum of Squares	Mean Square
	Intercept			
b	A	2	77.56	38.78
M	B	2	128.39	64.19
X	C	3	434.08	144.69
e	AB	4	36.28	9.07
X	AC	6	20.67	3.44
X	BC	6	75.17	12.53
X	ABC	12	50.83	4.24
	Residuals	0	0.000	

Effects Screen After Designating Effects for Analyzing Whole Plot Treatments

The designation of terms can be accomplished in numerous ways. As noted earlier they can be toggled from model to error (and back) by a double-click with your mouse. You've seen the right-click menu approach to designating terms, which provides a broader selection of option. Either click on the one you want, or enter its first letter. (No need to press the Alt key!) Finally, if you want to ignore a term, you can press the delete key. Go ahead and try all of the above, but be sure to return your selections to those shown above before proceeding.

Click on the ***ANOVA*** button. You can see that the whole plot treatment, pulp preparation (factor B), is significant. Ignore the remainder of the output for now, because this is only part of the picture.

Transform | Effects | ANOVA | Diagnostics | Model Graphs

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

Response: tensile in strength

ANOVA for Selected Factorial Model

Block term includes A

Error term includes AB

Analysis of variance table [Partial sum of squares]

Source	**Sum of Squares**	**DF**	**Mean Square**	**F Value**	**Prob > F**	
Block	77.56	2	38.78			
Model	128.39	2	64.19	7.08	0.0485	significant
B	*128.39*	*2*	*64.19*	*7.08*	*0.0485*	
Residual	36.28	4	9.07			
Cor Total	242.22	8				

ANOVA for Whole Plot

Do not look at the Diagnostics or the Model Graphs because the models are incomplete at this point.

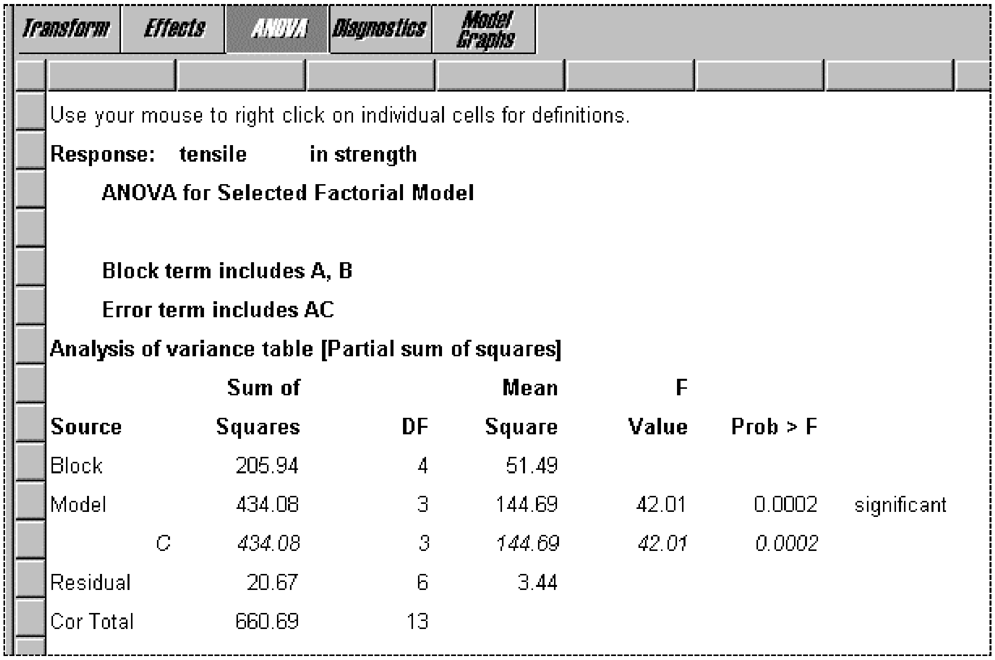
To analyze the subplot treatment (C - temperature), go back and click on the ***Effects*** button displayed in the toolbar at the top of the main window. The subplot treatment (C) should be tested against the Replicate by Temperature interaction, i.e., the AC interaction. Right click on factor **B** and choose **Block**, choose **C** for the **Model**, **AC** for the **Error** and **Ignore AB** and everything else. Your screen should now match the illustration shown below.

Transform | Effects

Term	DF	Sum
Intercept		
A	2	
B	2	
C	3	
AB	4	
AC	6	
BC	6	
ABC	12	
Residuals	0	

Designating Effects for Analyzing Subplot Treatments

Click on the ***ANOVA*** button. You can see that the subplot treatment, cook temperature (factor C), is significant. Ignore the remainder of the output.



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

Response: tensile in strength

ANOVA for Selected Factorial Model

Block term includes A, B

Error term includes AC

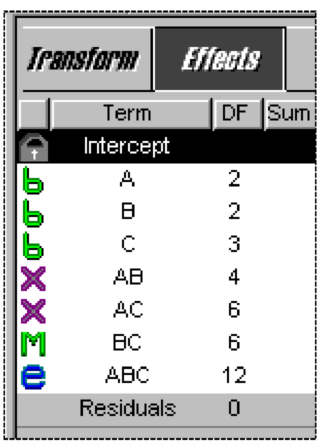
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	205.94	4	51.49			
Model	434.08	3	144.69	42.01	0.0002	significant
C	*434.08*	*3*	*144.69*	*42.01*	*0.0002*	
Residual	20.67	6	3.44			
Cor Total	660.69	13				

ANOVA for Sub-Plot

Do <u>not</u> look at the Diagnostics or the Model Graphs because the models are incomplete at this point.

To analyze the whole plot by subplot treatment interaction, click on the ***Effects*** button. The whole plot by subplot interaction (BC) should be tested against the ABC interaction. Right click on factor **C** and choose **Block**. Choose **BC** for the **Model**, **ABC** for the **Error** and **Ignore** the **AC** term. Your screen should now match the following illustration.



Term	DF	Sum
Intercept		
A	2	
B	2	
C	3	
AB	4	
AC	6	
BC	6	
ABC	12	
Residuals	0	

Designating Effects for Whole Plot by Sub-Plot Interaction

Click on the ***ANOVA*** button. The interaction in annotated view is labeled as insignificant. However, this is based on the Prob > F being over 0.05. Recall that earlier, the effect of B (the whole plot treatment) was deemed significant at a Prob > F of 0.0485. The Prob > F for BC is only slightly higher at 0.052. Therefore, the experimenters decided to ignore the advice from the software, and consider BC to be significant.

Transform | Effects | ANOVA | Diagnostics | Model Graphs

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

Response: tensile in strength

ANOVA for Selected Factorial Model

Block term includes A, B, C

Error term includes ABC

Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F	
Block	640.03	7	91.43			
Model	75.17	6	12.53	2.96	0.0520	not significant
BC	*75.17*	*6*	*12.53*	*2.96*	*0.0520*	
Residual	50.83	12	4.24			
Cor Total	766.03	25				

ANOVA for Whole Plot by Sub-Plot Interaction

Do <u>not</u> look at the ***Diagnostics*** or the ***Model Graphs*** because the models are incomplete at this point.

Before moving on to the final phase of the analysis, we want to provide some justification for the statistical analysis presented up to this point.

- In a blocked design the block by treatment interactions are used to estimate error.
- In a completely randomized blocked design all the block by treatment interactions are combined into a single estimate of error.
- Due to the restrictions on randomization in a split plot design we must separate out specific block by treatment interactions to estimate error for the particular treatment we are testing.
- To be able to separate out specific block by treatment interactions we enter the blocks (day) as factor (A).

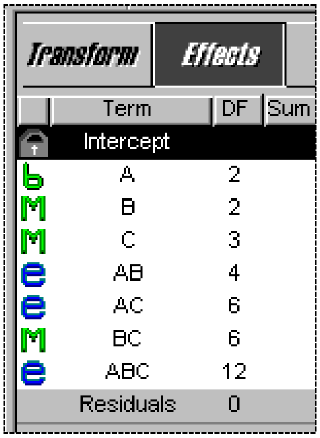
For this specific case, here’s how you did the tests for the ANOVA:

1. The whole plot treatment (B-pulp preparation) against the “block by pulp” (AB) interaction.

2. The sub-plot treatment (C-temperature) against the “block by temperature” (AC) interaction.

3. The pulp by temperature (BC) interaction against “block by pulp by temperature” (ABC) interaction.

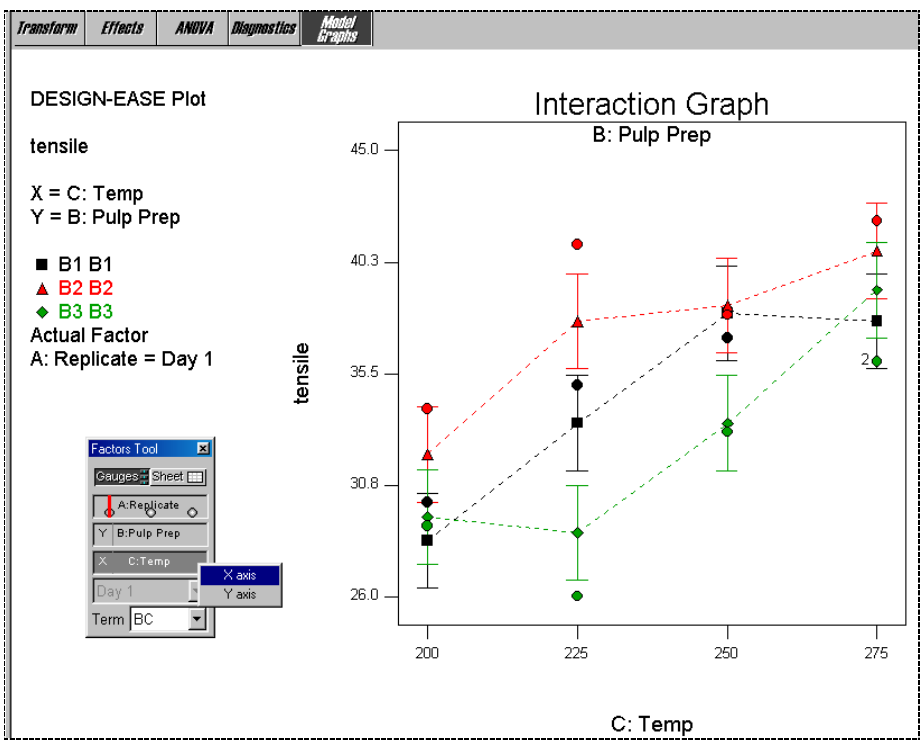
To get meaningful diagnostics and model graphs we need to fit the full model. Of course you must then ignore the ANOVA because the estimate of error will be incorrect. Click on ***Effects*** button displayed in the toolbar at the top of the main window. Choose **B, C** and **BC** for the **Model** and put **AB, AC** and **ABC** in the **Error**. Your screen should now match that shown below.



Fitting the Full Model for Purposes of Doing Diagnostics and Model Graphs

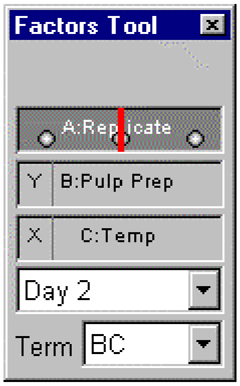
Click on the ***ANOVA*** button and ignore it. Click on the ***Diagnostics*** button and examine the diagnostic graphs. You should see nothing abnormal.

Click on the ***Model Graphs*** button. Right click over the **C: Temperature** factor on the tool palette and change it to the **X-Axis**. You now see an interaction plot of the continuous factor temperature at three discrete levels of pulp preparation.



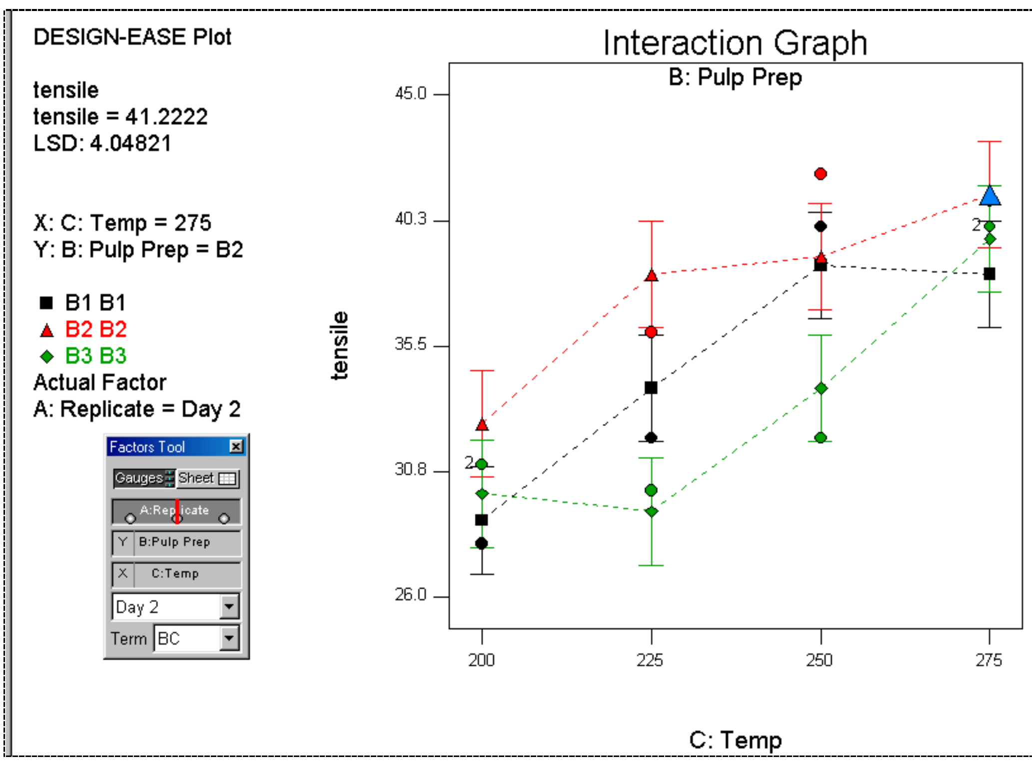
Effects Graph

By default, the graph comes up for Day 1. On the **Factors Tool** press the second button on **Replicate** the to see the predicted results for **Day 2**.



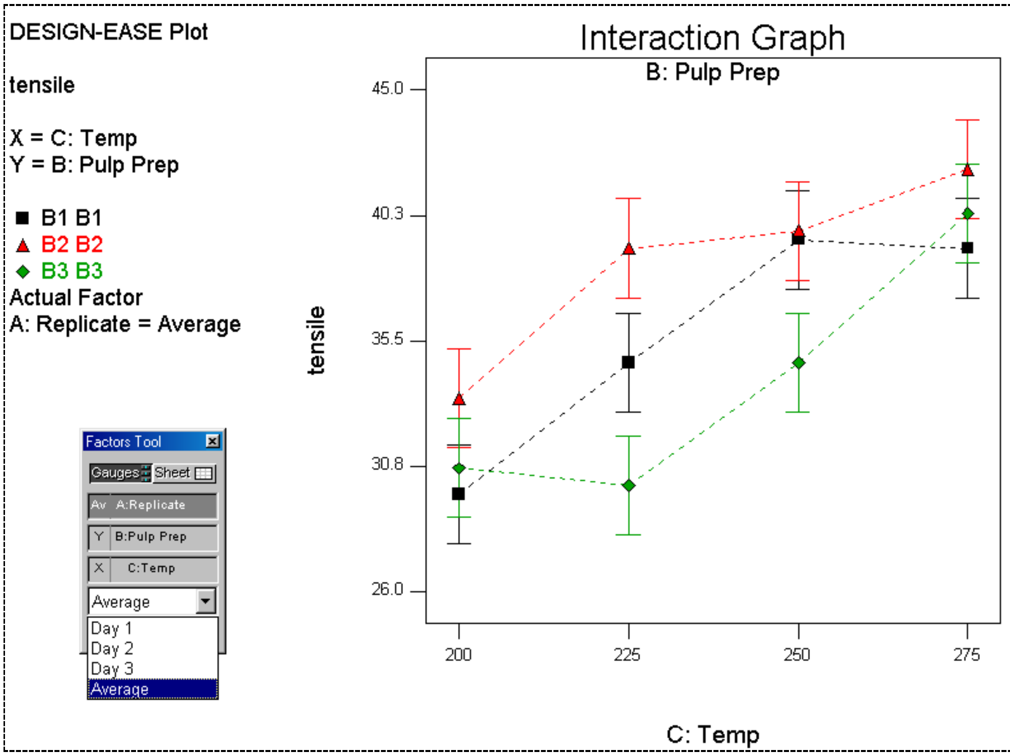
Setting Replicate to Day 2

Notice how the graph shifts. Click on any point to get the identification. The round symbols represent actual values, which are displayed to the left of the plot when you click the point. The other symbols (triangle, square and diamond) show the predicted means. Click any of these points to get the value for tensile strength and the least significant difference (LSD).



Interaction Graph for Day 2 with Top Tensile Prediction Clicked

Now press the third button to see the predicted results for **Day 3**. However, since time cannot be controlled, it might be best to show the BC interaction at an average of the Days. As shown below, click on the list arrow by the Day field and select **Average**.



Setting Replicates to Average Level

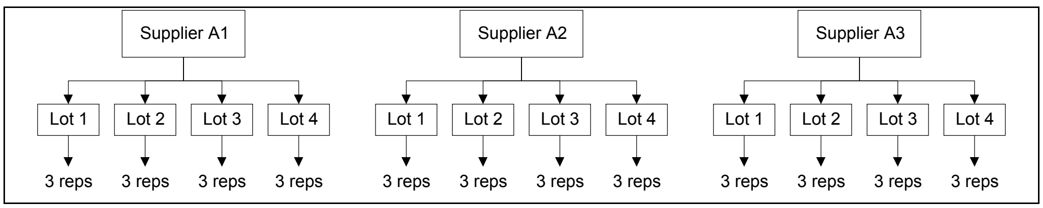
From this graph we can see pulp preparation B2 will maximize tensile strength and is most robust to temperature changes, especially in the temperature range of 225 to 275 degrees F. This relationship holds true regardless of which day (replicate) you select (the buttons on the tool palette).

Nested Design (Advanced Topic)

In some experiments the levels of one factor (e.g. factor B) are similar but not identical for different levels of another factor (e.g. factor A). This arrangement is call a “nested” or “hierarchical” design. The analysis of a nested design is tricky, even for statisticians. It can be done on Design-Ease by properly designating effects in specific ways for subsequent analysis of variance. Proceed with care!

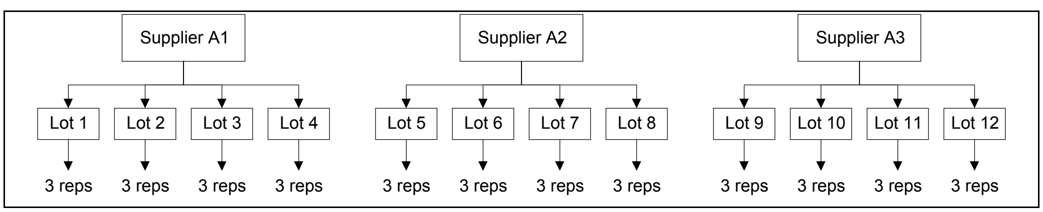
To illustrate how Design-Ease software can be used for a nested design, let’s do an example from Montgomery’s *Design and Analysis of Experiments*. A company buys raw material from three different suppliers. We want to know if raw material purity is dependent on the supplier. Four lots of raw material are selected at random from each of the suppliers. Three independent measures of purity are made on each batch of raw material.

As shown on the diagram below, this is a nested design.



Flowchart of Nested Design on Raw Material Supplier

The lots of raw material are nested within supplier. Lot 1 under supplier A1 is not the same lot as lot 1 under supplier A2 or A3. Another (perhaps more correct) diagram of this design is shown below. Note that the lots are unique.



A More Realistic Way to View the Lot-by-Lot Variation in Nested Design

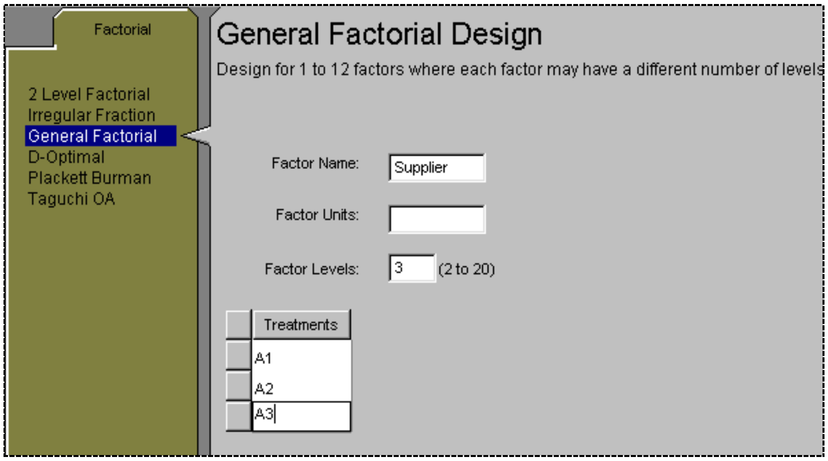
Here we can clearly see that the lot is dependent on supplier, i.e., lots of raw material are nested within supplier. Another name for a nested design is a hierarchical design. Think of the lots as children and supplier as a set of parents. Each lot is uniquely tied to its supplier, just as a child is to its parents.

Nested designs are a complex topic and this tutorial is intended to demonstrate only the mechanics. For theory see Montgomery’s *Design and Analysis of Experiments*.

Design the Experiment

Let's build the design. Click the blank-sheet icon 🗋 on the toolbar (or choose File, New Design). Then from the default **Factorial** tab click **General Factorial**. Enter **2** as the number of factors. Click on the **Continue** button.

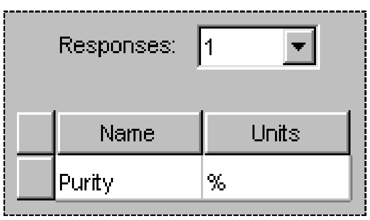
Enter **Supplier** for the factor name and **3** for the number of levels. Change the treatment names to "**A1**", "**A2**" and "**A3**". Your screen should now look like the illustration.



Entering Factor Names

Click on the **Continue** button to move on to the next factor. Then, enter **Lots** for the factor name, and **4** for the number of levels. Change the treatment names to "**B1**", "**B2**", "**B3**" and "**B4**". Click on **Continue**.

Enter **3** for Replicates (for the three independent measures of purity) and click on the **Continue** button. Leave the number of response at its default of 1. Enter the name as **Purity**, the units as **%**.



Response Names

Click on the **Continue** button. Design-Ease now displays the experiments (samples to test) in random run order.

Analyze the Results

Read in the response data via **File**, **Open Design** from the main menu. Select the file named **Purity.de6**. The purity values are based on a nominal value of 93 percent. For example, a value of 4 indicates a purity of 97 (93 + 4). Presumably this rescaling of the response made it easier for people to interpret the results.

Since this is a nested design, the analysis is not as straightforward as usual. (Warning: If you're not a statistician, you may want to skip to the next highlighted command, and not look at the table or accompanying detail on mean squares (MS), etc., etc.!) In order to perform the correct F-test, you must determine which terms belong in the model and in the error. (Remember that $F=MS_{model}/MS_{error}$.)

For our example, B is a random factor (a sample from a population of lots) that is nested within A, which is a fixed factor (there are only three suppliers of interest). The expected mean squares (EMS) are shown in the table below.

Source	**df**	**Expected MS**	**Calculated MS**
A - Supplier	2	$\sigma^2 + 3\sigma^2_B + 6\Sigma A^2$	$(SS_A)/(df_A)$
B - Lots	9	$\sigma^2 + 3\sigma^2_B$	$(SS_B + SS_{AB})/(df_B + df_{AB})$
Error	24	σ^2	$MS_{Pure\ Error}$
Total	35		

Expected Mean Squares

See Montgomery's *Design and Analysis of Experiments* chapters 11 and 12 for details on expected mean squares.

From the expected MS column it can be inferred that A (Supplier) should be tested against the nested factor, B (Lots). The appropriate test for significance of A is then:

$$F = (\sigma^2 + 3\sigma^2_B + 6\Sigma A^2)/(\sigma^2 + 3\sigma^2_B).$$

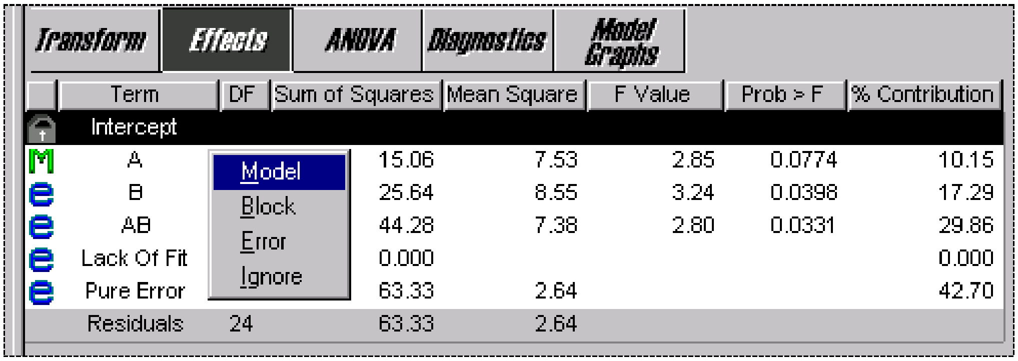
From the calculated MS column it can be seen that the correct sum-of-squares for B is created by adding the sum-of-squares for factor B and the AB interaction. You must add these sums of squares manually using the procedure outlined below. Since lots are nested within supplier, there can be no true supplier by lot interaction (AB). Therefore the sum of squares (SS) is:

$$SS_{(B\ within\ A)} = SS_B + SS_{AB}.$$

To analyze factor A (supplier), click on the analysis node labeled **Purity**, which can be found in the tree structure along the left of the main window. Then, click on the ***Effects*** button displayed in the toolbar at the top of the main window. There are four states an effect can have:

- Model ("M")
- Block ("b")
- Error ("e")
- Ignore ("X").

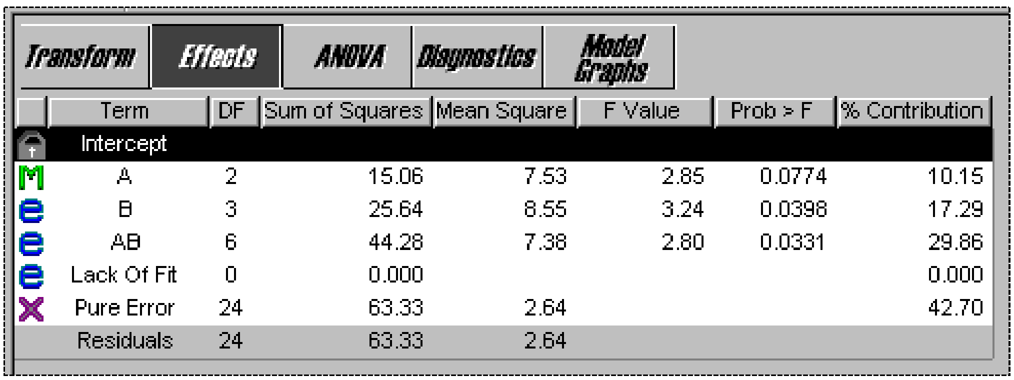
Supplier (A) is tested against the lots within supplier (B + AB). Right click on factor **A** and choose **Model**.



Term	DF	Sum of Squares	Mean Square	F Value	Prob > F	% Contribution
Intercept						
A		15.06	7.53	2.85	0.0774	10.15
B		25.64	8.55	3.24	0.0398	17.29
AB		44.28	7.38	2.80	0.0331	29.86
Lack Of Fit		0.000				0.000
Pure Error		63.33	2.64			42.70
Residuals	24	63.33	2.64			

Designating the Effect of A for Model

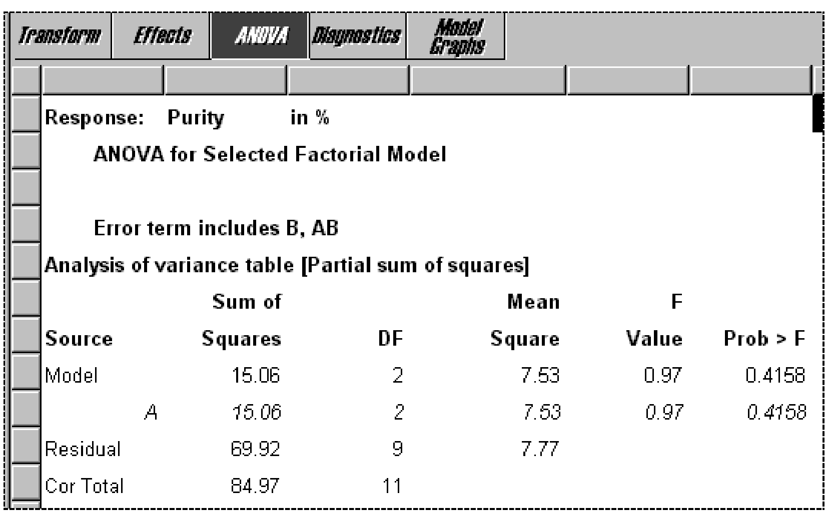
Right click on factor **Pure Error** and choose **Ignore**. Set (or leave) **B** and **AB** and **Lack of Fit** as **Error**. Your screen should now match the illustration below.



Term	DF	Sum of Squares	Mean Square	F Value	Prob > F	% Contribution
Intercept						
A	2	15.06	7.53	2.85	0.0774	10.15
B	3	25.64	8.55	3.24	0.0398	17.29
AB	6	44.28	7.38	2.80	0.0331	29.86
Lack Of Fit	0	0.000				0.000
Pure Error	24	63.33	2.64			42.70
Residuals	24	63.33	2.64			

Completed Effects Screen in Preparation for ANOVA on Effect of A

Click on the ***ANOVA*** button. You can see that the supplier (factor A) is not significant (Prob > F is greater than 0.10). Don't bother looking any further at this report.



Response: Purity in %

ANOVA for Selected Factorial Model

Error term includes B, AB

Analysis of variance table [Partial sum of squares]

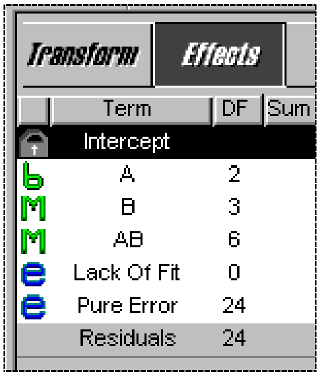
Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	15.06	2	7.53	0.97	0.4158
A	*15.06*	*2*	*7.53*	*0.97*	*0.4158*
Residual	69.92	9	7.77		
Cor Total	84.97	11			

ANOVA for Effect of A

Do <u>not</u> look at the Diagnostics or the Model Graphs because the model is incomplete at this point.

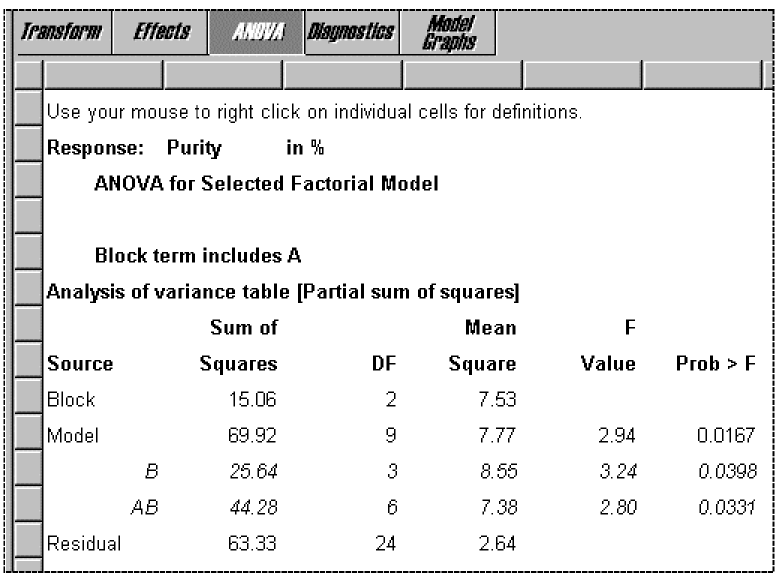
To analyze the nested factor (B-lots), go back and click on the ***Effects*** button displayed in the toolbar at the top of the main window.

The lots within supplier should be tested against the pure error. Set this up by right-clicking on the designations for the various terms as follows: **Block** on **A**, **Model** for **B** and **AB**, and **Error** for **Pure Error**. Your screen should now match the illustration below.



Designating Effects for ANOVA

Click on the **ANOVA** button. The results show that lots within supplier (the Model line in the ANOVA) are significant.



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

Response: Purity in %

ANOVA for Selected Factorial Model

Block term includes A

Analysis of variance table [Partial sum of squares]

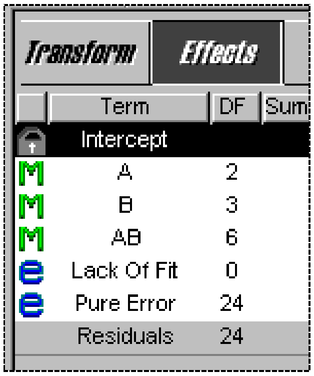
Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Block	15.06	2	7.53		
Model	69.92	9	7.77	2.94	0.0167
B	*25.64*	*3*	*8.55*	*3.24*	*0.0398*
AB	*44.28*	*6*	*7.38*	*2.80*	*0.0331*
Residual	63.33	24	2.64		

ANOVA for Lots Within Supplier

Do not look at the Diagnostics or the Model Graphs because the model is incomplete at this point.

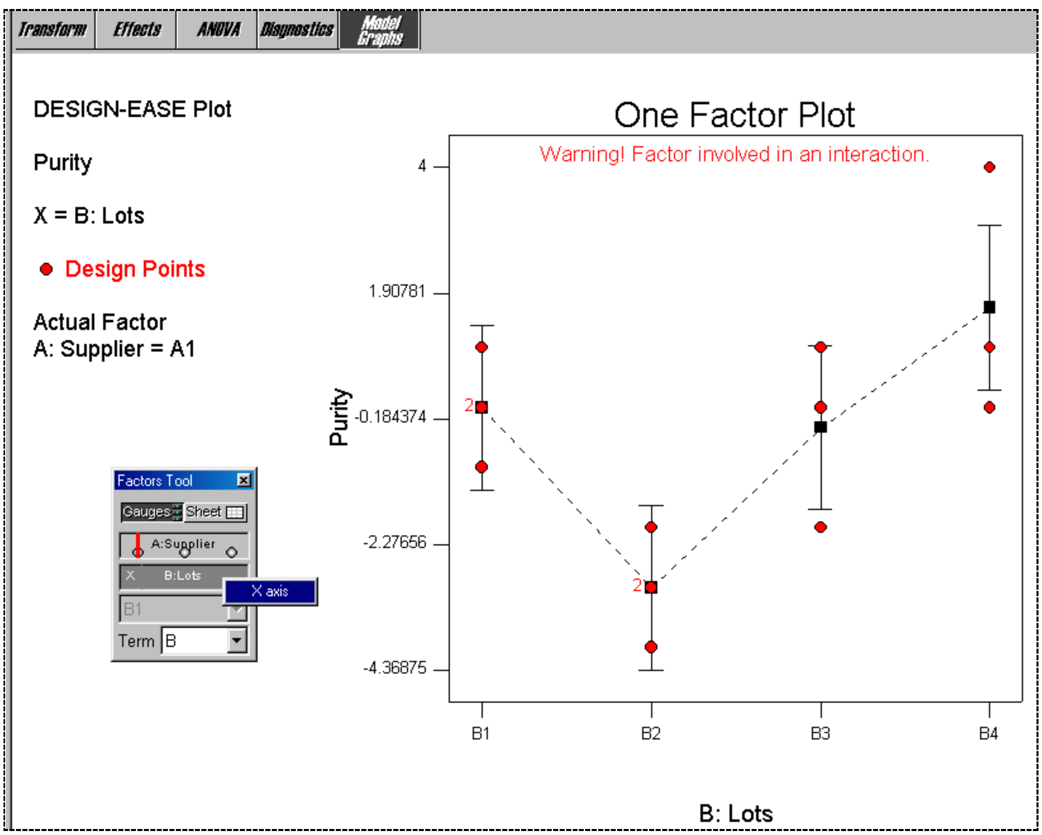
To get meaningful diagnostics and model graphs you need to fit the full model. Of course you must then ignore the ANOVA, because the estimate of error will be incorrect.

Click on ***Effects*** button displayed in the toolbar at the top of the main window. By right-clicking, choose **Model** for **A**, **B**, and **AB** and **Error** for the **Pure Error** and **Lack of Fit**. Your screen should now match that illustrated.



	Term	DF	Sum
	Intercept		
M	A	2	
M	B	3	
M	AB	6	
e	Lack Of Fit	0	
e	Pure Error	24	
	Residuals	24	

Click on the ***ANOVA*** button and ignore it. Click on the ***Diagnostics*** button and examine the diagnostic graphs. They look good, so click on the ***Model Graphs*** button. From the main menu bar choose **View**, **One Factor**. The default graph shows the effect of supplier on purity, but this is not statistically significant, so right click over the **Lots** factor on the floating graph tool and change it to the **X-Axis**. Lots within supplier are significant. (Ignore the warning shown on the graph.)



Graph of Purity versus Lots for Supplier A1

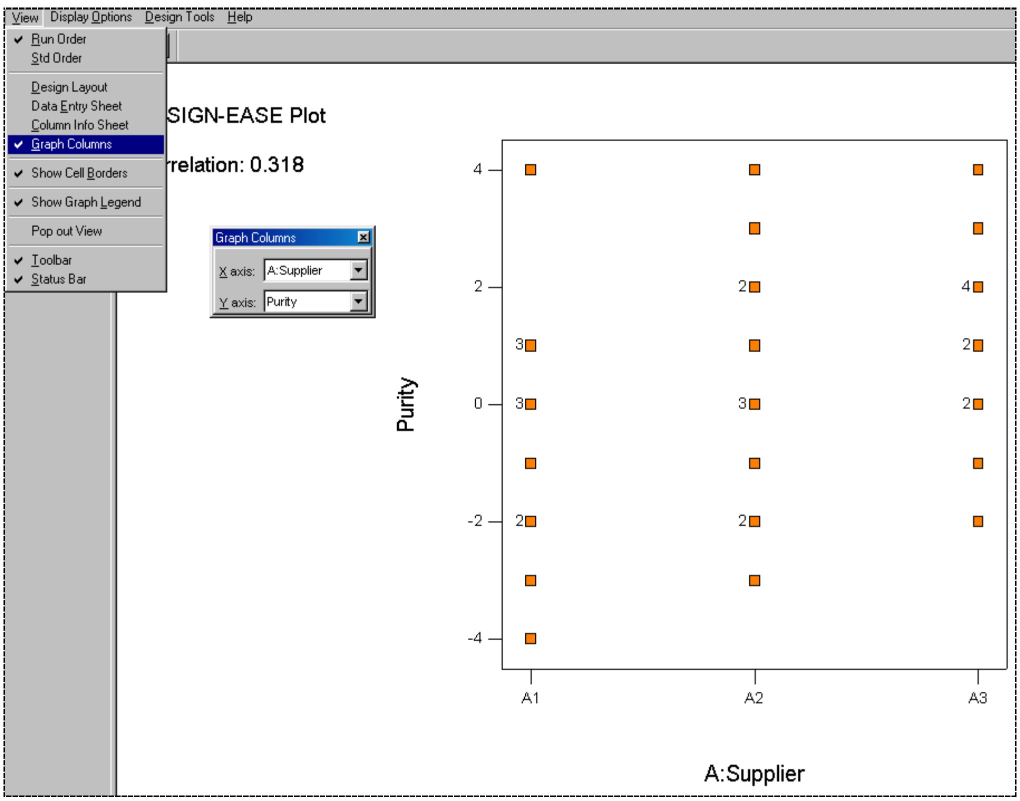
Note: the number 2's shown for lots B1 and B2 indicate that there's an actual point (round symbol) hidden by the mean results (square symbols).

Click on the other two buttons for **A:Supplier** to see how lots vary within suppliers **A2** and **A3**. The patterns change in random fashion, but the amount of variation is considerable.

Montgomery summarizes this case study by saying:

"*The practical implications of this experiment and the analysis are very important. The objective of the experimenter is to find the source of the variability in raw material purity. If it results from differences among suppliers, then we may be able to solve the problem by selecting the "best" supplier. However, that solution is not applicable here because the major source of variability is the batch-to-batch purity variation within suppliers. Therefore, we must attack the problem by working with the suppliers to reduce their batch-to-batch variability.*"

To provide support for Montgomery's conclusions, click on the **Design** node. Then select **View**, **Graph Columns** to see a scatterplot of purity versus supplier.



Scatterplot of Purity versus Supplier

(Note: the numbers by some points indicate repetitive values. For example, there were three purity results for supplier A1 that fell at zero – the nominal value of purity.) The correlation of purity versus supplier is positive in the direction of supplier A3, but only very weakly. There's too much lot-to-lot scatter within each supplier to see statistically significant differences between them.