# Lab 13: Factor Screening and Fractional Factorial Design (Chapter 7)

## Objectives

* Create fractional factorial designs in JMP
* Learn how to translate between real values and coded values in order to “run” actual experiments
* Use the stepwise function in JMP to screen factors and rule out which are significant

Up to now, our data sets generally had more than a single replicate for each experiment but very few factors. Therefore, we could simply run a replicate of the design of experiments to get more data to estimate the error. Even with 3 factors, this only requires an additional 23=8 runs. But when k becomes 4 or larger, this number of runs starts to increase quite rapidly. Economics of time and money will usually prevent us from performing or repeating a full factorial design. There is a principle called sparsity of effects that tells us the systems we study are usually dominated by first and perhaps low-order interactions. Typically, we use this principle to exclude 3rd order and higher terms from the regression model. For example, if we are considering factors A, B, C, D, the model may just be dominated by the terms that are A, B, C, D, or A\*B, B\*C, etc. and not A\*B\*C, B\*C\*D, etc. **There are two ways we can reduce the number of experiments we run even more: reduce the number of runs via fractional factorial experimental designs or screen our factors and remove non-significant ones.**

One way to reduce the number of experimental runs we do is to perform a fractional factorial design of experiments. We have already established that most higher order interaction terms (order 3 and above) will never be significant. This allows us to run a single replicate design and leave those higher order terms out of the model to get an estimate of the variability. But in order to reduce the number of required runs further, we now explore a technique which sacrifices the resolution of effects even more by mathematically lumping certain effects together.

We may find that not all of the factors in a model are significant and thus do not need to be included in the model at all. **Therefore, as a second way to reduce the number of experimental runs you need to perform, you will want to first screen for significant factors then remove unneeded insignificant parameters from your models in order to allocate more resources toward getting a better estimate for your model**. This will not always guarantee a lower variance in the model, since less parameters are being fit and therefore less of the total error will be explained by the model. However, if we repeat runs, this allows us to run a smaller 2k design; assuming some factor and any interactions it is involved in are insignificant, it can be dropped effectively collapsing the original 2k design onto a smaller design, doubling the number of replicate points for each factor dropped from the design. In general, if we delete h factors so that r = k - h factors remain, the original 2kdesign with n replicates will project into a 2r design with n \* 2h replicates.For example, with our factors A, B, C, and D, say after analyzing the data from the 16 runs, it was determined that factor D was not significant. So by removing D from the design of experiments, the 16 runs now can be considered 2 replicates of 8 runs for a three factor full factorial design. If C was also removed, then the 16 runs could be considered 4 replicates of a full factorial design on A and B. Since C and D are not significant, the high and low settings don't matter in this reduced design.

## Fractional Factorial Design

When creating a screening design, instead of creating a design that includes every experimental run, you can choose fractional factorial designs to reduce your number of runs. After adding your factors and clicking Continue, you will have the option of choosing between several different designs. **For this class, you don’t need to consider any design that involves blocks.** This leaves 3 options: 32 runs for a full factorial, 16 runs for a 1/2 fraction factorial (Resolution V), and 8 runs for a 1/4 fraction factorial (Resolution III). Let's start by choosing the 1/2 fraction with 16 runs.

## 

You can also add center points (generally, 2-4) to help make your model even better (but remember that adding 2-4 center points adds 2-4 runs).

## Translating Coded/Real Values

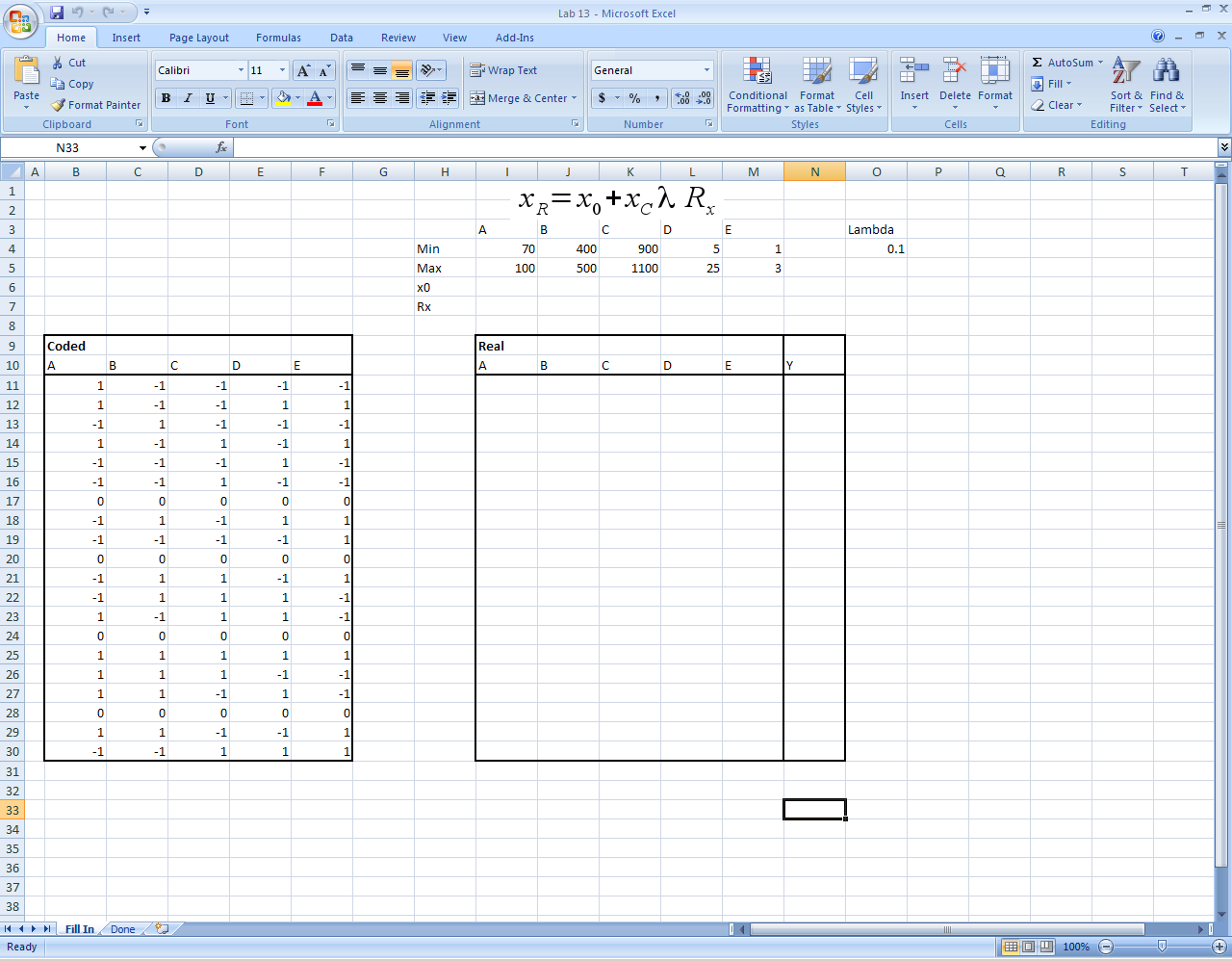
## After you’ve created your DOE in JMP, you have a set of coded values (-1 and 1) to run your experiments. However, in the real world, your high and low values won’t be -1 and 1, so we must use the following equation to translate coded values to real values.



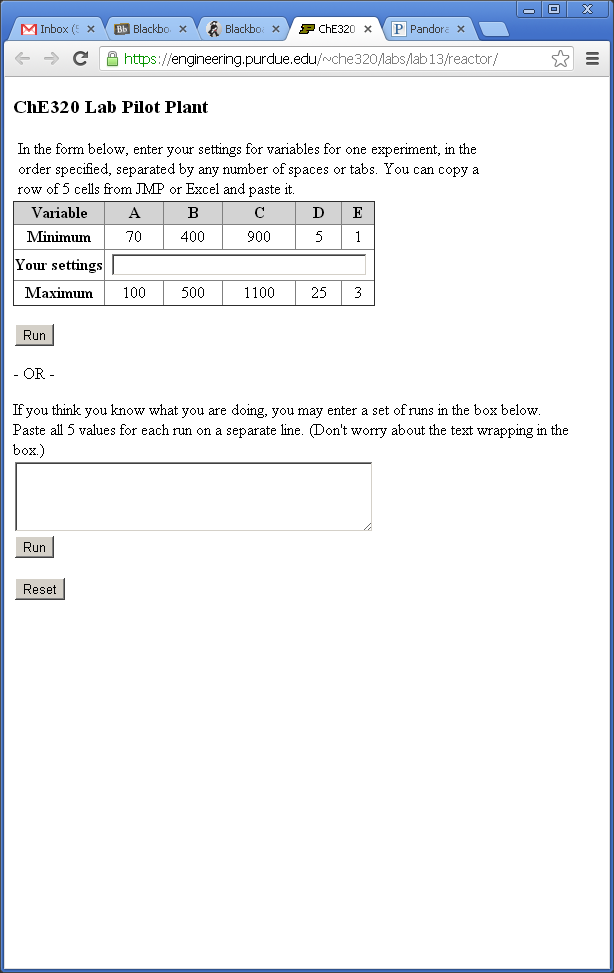
xC is the coded value. xR is the real value. x0 is the value of X at the center of the design. Rx is the potential range of X. λ is a scaling factor to reduce the size of the screening design to some percent of the full range of X.

A λ of 0.5 will cover the entire range of X if the design is already centered in the middle of the factor. (Remember, xC goes from -1 to +1, a coded range of 2, so 0.5 \* 2 covers the entire range of the real valued X. **When we are screening for significant effects, we are forming a linear model. The larger the real distance between the +1 and -1 levels of a factor, the greater the chance that our linear model will not give satisfactory results.** We typically wish to have the size of the design to be as small as possible while still verifying that some change is the response occurs, so a lambda value of .05-.1 is best.

It’s easiest to do this translation in Excel, like in the spreadsheet below.



After you’ve decoded your values, you can run your experiments with these inputs; or, in the case of this class, input these numbers into the website simulator (<https://engineering.purdue.edu/~che320/labs/lab13/reactor/>).



After you’ve obtained your outputs, insert your Y values back into JMP, then continuing to screen for significant factors, create your model, and analyze the output.

## Using JMP to Screen for Significant Factors

## After opening your data in JMP, choose Analyze > Fit Model. Add your response and factors and select Macros > Factorial to Degree.

## 

## Also, change the Personality to Stepwise.

## 

## Click Run Model. From here, you can add terms into the model in order of significance. Clicking Step adds in a factor, one at a time, until you’ve added all terms that have a P-value less than the Prob to Enter value (which is your alpha value, usually .05). You can also click Go to add in all of the terms at once. Conversely, if you set Direction to Backward, you can add in all of the parameters at once (Enter All) then click “Step” to remove the non-significant ones one-by-one.

## 

## Once you’ve entered all of your significant terms (or removed all of the non-significant terms), you can click Make Model to create your regression model.

## Summary of DOE so far…

Today’s lab expands on last week’s lab on DOE. To summarize, these are the steps we know thus far:

1. Create a DOE in JMP, either full factorial if we have <3 factors (lab 12) or fractional factorial if we have >3 factors (lab 13)
2. Obtain your experimental data by:
   1. Translating your coded values outputted by JMP into real values
   2. Running those real values as experiments (or, in this class, in a website simulator)
3. Screen for significant effects using stepwise regression (lab 13)
4. Analyze your model given the outputs from JMP (lab 12)

## Lab 13 Exercises

Use the simulator website (<https://engineering.purdue.edu/~che320/labs/lab13/reactor/>) as your data source for the lab. Screen at the center of the range for each factor using a λ value of 5% or 10%.

* Use a JMP to design a 1/2 fraction (16 runs) factorial design of experiments for the 5 factors in the lab pilot plant website.
* Copy the design to a spreadsheet with any other useful information from the website.
* Translate the coded design of experiments to actual values.
* Copy a single run and paste into the website. Then hit the run button to see the output.
* Copy the entire table of runs into the lower box on the website. Hit the run button. Now copy the results back into the spreadsheet and JMP.
* Analyze the results in JMP

1. Which parameters are significant?
2. Do the center points indicate a linear model is OK to be assumed?
3. Find all the variance estimates in the final model output. Are they similar?
4. Check the residuals. Is anything suspect?
5. What is the final fraction/replicate count after reducing the model to only the significant terms?

Feel free to repeat this exercise until you are very comfortable with the procedure. Try using different λ values and different fractional or full factorial designs. This is the same interface and procedure you should be using on the first part of your project.