STAT COE-Report-02-2018

Classical Designs: Fractional Factorial Designs

Authored by: Cory Natoli

30 June 2018

Revised 30 September 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 30 Sep 2018: Minor formatting edits.

Executive Summary

When designing a test, one common approach is to use a classical design. Classical designs include full factorial designs and fractional factorial designs. This paper covers when and how to use fractional factorial designs and assumes knowledge of full factorial designs (Montgomery 2017). Fractional factorial designs are very useful for screening experiments or when sample sizes are limited. However, some information gained from a full factorial design can be lost when using a fractional factorial design. It is important to understand what these drawbacks may be and when the risk associated with them is acceptable. Step-by-step instructions are provided on how to create these designs in the statistical software package JMP.

Keywords: Screening, interactions, main effects, Fractional Factorial designs, test, JMP

Introduction

Full factorial designs are a common starting point when planning a test, but as the number of factors becomes large, the size of the design grows very quickly. If we assume each factor has two levels, a full factorial design (called a 2^k design) with 8 factors would require 256 (2^8) runs! In many tests, some categorical factors have more than 2 levels, which further increases the test size of a full factorial. However, resources may be limited and it is typically impractical to conduct a test of this size. As the number of factors increases, the 2^k full factorial design grows prohibitively large. Fractionating the factorial design sacrifices information about some of the interactions in favor of reducing the total number of runs. If your motivation is to identify the vital few (significant) factors from the trivial many (screening), then a fractional factorial design is an efficient alternative to a full factorial design. This best practice shows why and when fractional factorial designs are useful, as well as the risk associated with using a fractional factorial design.

When to use a fractional factorial design

A fractional factorial design is a reduced version of the full factorial design, meaning only a fraction of the runs are used. A fractional factorial design allows for a more efficient use of resources as it reduces the sample size of a test, but it comes with a tradeoff in information. The main use for fractional factorial designs is in screening experiments. Montgomery (2017) describes screening experiments as "tests in which many factors are considered and the objective is to identify those factors that have significant effects." He also adds that "screening experiments are usually performed in the early stages of a project when many of the factors initially considered likely have little or no effect on the response." The goal of a screening experiment is to get an initial sense of what factors are important and need to be studied further. This naturally leads to a sequential testing strategy in which follow-on testing can be done after the first fractional factorial test. Fractional factorial designs allow us to reduce the number of runs needed in an initial test without sacrificing too much loss of information. However, it is important to remember that there will be some risk associated with reducing the number of runs.

In cases where a test has factors with more than 2 levels, constrained test spaces, unusual run size restrictions, or strict time constraints, fractional factorial designs are not recommended. Generally, a computer-generated optimal design can more effectively handle these types of scenarios. Another important consideration when creating a fractional factorial design is the model that you wish to estimate after the test is completed. If you wish to estimate all terms (including the complex high-order interactions) and can only execute a test one time (as opposed to a sequential approach), a fractional factorial design may not be the best option. This scenario is discussed further in later sections.

Creating a fractional factorial design

First consider a simple case with three factors to create a one-half fraction. For example, a tester may want to know if range, altitude, and airspeed affect miss distance of a weapon system. Assume we wish to design a test that includes all three factors with two levels each. The full factorial requires 8 (2^3) runs and is shown in Table 1 where the entries are in coded units so that the "-" denotes the low level of the factor and the "+" denotes the high level of the factor.

Table 1: Full Factorial Design Matrix for a three 2-level Factor Test

Run	I	A	B	C
1	+	+	+	+
2	+	+	+	-
3	+	+	-	+
4	+	+	-	-
5	+	-	+	+
6	+	-	+	-
7	+	-	-	+
8	+	-	-	-

The "I" column in Table 1 denotes the intercept of the model and is always a column of +'s. When creating a linear regression model, the form of that model is:

$$Y = \beta + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3$$

We include a term for each factor, interaction, and the intercept. This is why the "I" column exists in the design matrix. This column can also be used with "generating equations" when building the fractional factorials manually/theoretically. This paper does not cover this theoretical side of fractional factorial designs as we will use software to create these designs. More information can be found in Montgomery

(2017). The remaining columns denote the levels that the factors are set at for the 8-run test to be conducted. We can extend the design matrix into the model matrix (Table 2), which includes columns representing the interactions. An interaction term in the model explains the effect of one factor on the response at different values of another factor. In the model matrix, the interaction is simply the multiplication of the levels of the factors included in the interaction (e.g., AB for run 1 is +, because $+ \times + = +$, or in equivalent notation, $A \times B = AB$).

Table 2: Full Factorial Model Matrix for a three 2-level Factor Test

Run	I	A	B	C	AB	AC	BC	ABC
1	+	+	+	+	+	+	+	+
2	+	+	+	-	+	-	-	-
3	+	+	-	+	-	+	-	-
4	+	+	-	-	-	-	+	+
5	+	-	+	+	-	-	+	-
6	+	-	+	-	-	+	-	+
7	+	-	-	+	+	-	-	+
8	+	-	-	-	+	+	+	-

The full factorial design allows us to estimate each of these terms: the intercept, main effects, two-factor interactions, and even the three-factor interaction. This property extends for more than three factors. A full factorial design allows you to estimate all interaction effects from the two-factor interaction through the k-factor interaction. To create a fractional factorial design, we need to strategically reduce the number of runs in the full factorial design in half. By definition, a full factorial design can be divided into two half fractions: a principal fraction (the rows highlighted in yellow in Table 3) and an alternate fraction (the rows not highlighted). Which fraction to run is not critical unless one fraction is particularly difficult to run or one fraction contains a test combination of particular interest to the experimenters/subject matter experts. The information in either fraction is statistically equivalent.

Table 3: Fractional Factorial Model Matrix for a 3 2-level Factor Test (yellow highlighted rows make up the principal fraction)

Run	I	A	B	C	AB	AC	BC	ABC
1	+	+	+	+	+	+	+	+
2	+	+	+	-	+	-	-	-
3	+	+	-	+	-	+	-	-
4	+	+	-	-	-	-	+	+
5	+	-	+	+	-	-	+	-
6	+	-	+	-	-	+	-	+
7	+	-	-	+	+	-	-	+
8	+	-	-	-	+	+	+	-

The method to accomplish this is to confound the highest order interaction (ABC for this three factor example) with the fraction. That is, the principal and alternate fractions are determined by grouping runs together where the highest-order interactions have like signs. In Table 3, the principal fraction (highlighted in yellow) contains all of the rows with a value of "+" for the ABC interaction, and the alternate fraction contains the remaining runs that have the "-" ABC values. This is called a half-fraction factorial design and is denoted as a 2^{3-1} design. When you actually perform the test, ensure the runs are executed in a random order.

By executing a fraction of the runs of a full factorial design, aliasing in model effects is introduced. Two model terms (e.g., main effect and two-factor interaction) are aliased when their columns in the model matrix are identical for every row. If we look at the principal fraction for the 2^{3-1} design in Table 3, we can see that main effect A is aliased with BC, main effect B is aliased with AC, and main effect C is aliased with AB. Because of this aliasing, when analyzing the results of the half-fraction test, the researcher cannot differentiate between the effects that are aliased with each other. If, for example, A was determined to be a statistically significant factor, the researcher would not be able to say whether factor A or the interaction BC was causing an impact on the response. We explore the impact of aliasing on our designs throughout this paper. This technique to generate a half-fraction can be done for any number of factors: take the highest order interaction term and split the rows by the values in that column. Smaller fractional factorial designs also exist outside of just the simple one-half fraction. In order to generate this, more than one high order interaction will have to be used to divide the runs into the fractions. Statistical software is able to choose which interactions are used that minimize overall aliasing. As the number of factors increases, it is not unusual to see one-quarter fractions or one-eighth fractions.

Aliasing and Resolution

In order to visualize the aliasing in a fractional factorial design, you can examine a color map of correlations. Figure 1 is an example of a color map created in JMP for a design with 8 factors.

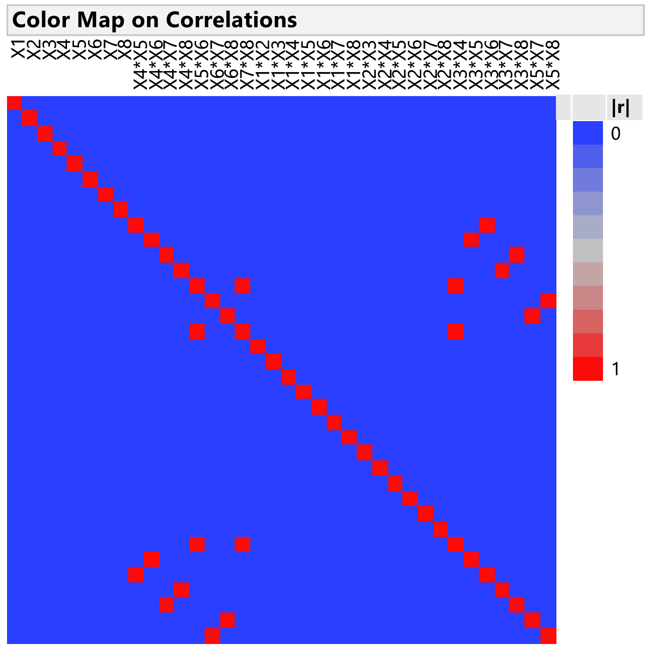


Figure 1: Color Map on Correlations in JMP

A color map of correlations shows the absolute value of the correlation, r, between two model terms (e.g. a main effect or interactions). The matrix is diagonal so that each row and column represents the main effects and interaction terms in the model. The correlation values can range from 0 to 1 in magnitude. The higher this value, the more correlated the terms are, and the harder it is for the analyst to differentiate between the effect and the effect(s) it is correlated with. When two terms are correlated, we say these terms are aliased or confounded. When the magnitude of the correlation between two terms is equal to 1, then we say those terms are perfectly aliased or completely confounded. In the JMP output, red squares indicate perfectly aliased terms. Ideally, a color map of correlations would have red squares on the diagonal, as all terms are perfectly correlated with themselves, and blue squares in the off-diagonal (this is what we would expect to see when doing a full factorial and include all higher order interaction terms). In a fractional factorial design, we must be willing to accept the risk of aliasing in order to save runs (and money). This affects our ability to analyze the data after the experiment has been conducted and may limit decision-making capabilities. A sequential testing strategy can help resolve these issues as needed.

When creating a fractional factorial design, the aliasing scheme determines the resolution of the design:

- **Resolution III**: Main effects are aliased with two-factor interactions.
- **Resolution IV**: Main effects are not aliased with two-factor interactions; however some two-factor interactions are aliased with other two-factor interactions.
- **Resolution V**: Main effects are not aliased with two-factor interactions; furthermore, two-factor interactions are not aliased with other two-factor interactions.

The lower the resolution, the less information that can be gained from the test. The JMP output in Figure 1 is an example of a Resolution IV design as two-factor interactions are aliased with other two-factor interactions, but none of the main factors are aliased with each other or with any two-factor interactions. The 2^{3-1} design in Table 3 is a Resolution III design since main effects are aliased with two-factor interactions. Resolution III designs should never be run in practice.

Figure 2 shows the primary areas of concern to investigate on a broken out color map. We are not concerned with interactions beyond the two-factor interactions due to the sparsity of effects principle that is covered in the next section.

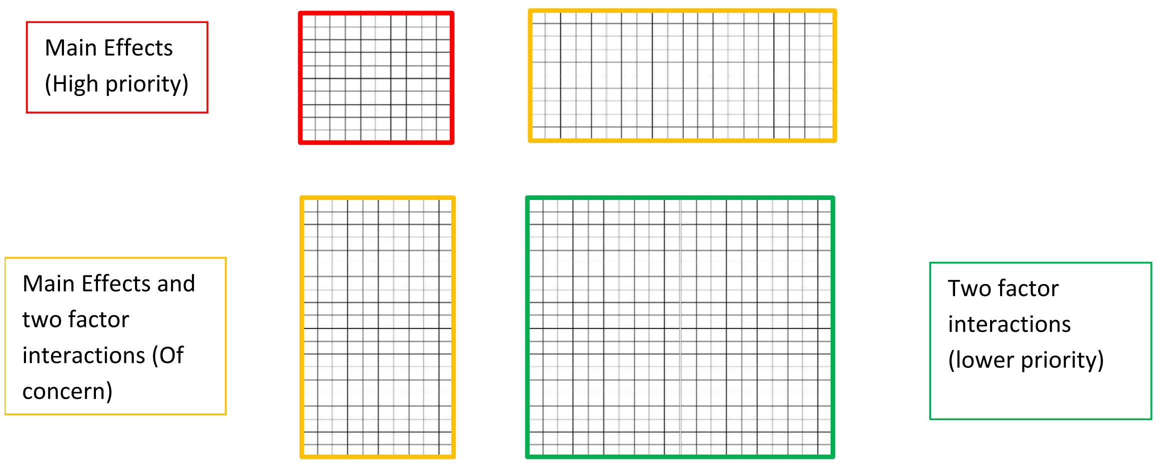


Figure 2: Areas of importance in a Color map on Correlations.
(Adapted from Errore, Jones, Li, & Nachtsheim, 2017)

The highest priority is to make sure that main effects are not aliased with one another and with two-factor interactions. Since fractional factorial designs are generally used for screening, it is crucial that the researcher can distinguish which main effects are statistically significant. However, beyond the main effects there will be a tradeoff between the number of runs and information gained. The objectives in screening designs are less concerned about learning everything about a system, and focus on identifying important factors. Therefore, researchers may be willing to accept the risk of determining an aliased factor is significant. If in reality a significant factor is present, it will be impossible to discern which of the aliased terms is truly statistically significant from this phase of testing. Therefore, it is important to understand the aliasing structure before testing to determine which risks will be acceptable. While using a Resolution IV design can be useful in certain situations, choosing a design with two factor interactions aliased with other two factor interactions can be risky and for this reason Resolution V designs are the gold standard.

Fractional factorial design properties

Full factorial designs allow us to estimate the effect and determine the significance of all interactions within a test. For example, if we had eight factors of interest, we would be able to estimate the 8-factor interaction after conducting a full factorial design. However, the sparsity of effects principle states that typically only a few factors of many are important in a model. A system is often dominated by main effects and two-factor interactions. Three-factor interactions may be seen occasionally, but four-factor interactions and above are very rare (Montgomery, 2017). While subject matter expertise may suggest that three-factor or higher interactions cause an effect on the response, this principle tells us that the main effects and two-factor interactions will typically satisfactorily describe the overwhelming majority of variability in the response in our statistical model. Remember, the goal of this model is to identify the key factors and interactions that cause an effect on our response and the higher-order terms are generally minor (or not statistically significant) effects. Knowing this property, we can choose to accept the risk of neglecting three-factor interactions and above in order to save resources (runs) by running a fractional factorial design.

The projection property states that fractional factorial designs 'project' into either replicated or factorial designs, higher resolution fractional factorial designs, or full factorial designs in fewer factors if insignificant effects are removed from the model. For example, if we have a 3-factor test and determine that factor A is insignificant, the design will project into a full factorial for factors B and C. This allows us to estimate all model terms with B and C, as well as eliminating any potential aliasing problems that arose from the initial design. In some situations, this property allows for the use of lower resolution fractional factorial designs when subject matter experts anticipate one or more factors in a test will not be significant. It also helps to obtain a better estimate of noise in the system on the factors that are significant. Figure 3 shows how a fractional factorial design might project if insignificant factors are found.

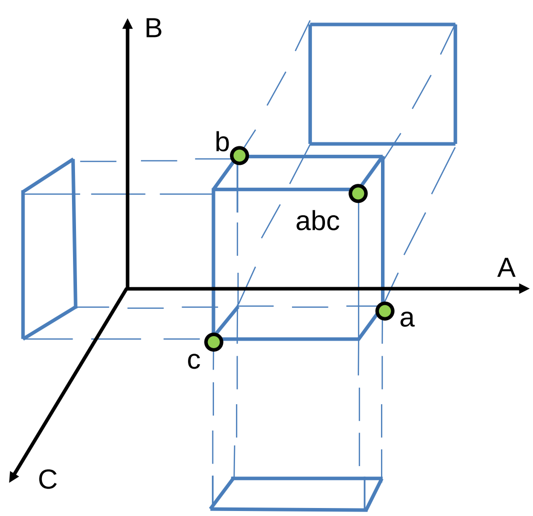


Figure 3: Projection property for fractional factorial designs

Possible outcomes after executing a fractional factorial design

Another key to success for using fractional-factorial designs is to understand the concept of sequential experimentation. If a fraction does not allow you to determine which effects are significant because of aliasing, run an additional fraction by folding over the original design (fold over designs are covered later in this paper). The first set of tests should include only a portion of the total number of tests you plan to run:

- Run the initial set of tests.
- Change factors, levels, ranges, or aliasing as needed in subsequent sets of tests; the need is based on information gained from the previous set of tests.
- Conduct confirmation runs to validate your findings.

This method of developing a new set of tests based upon knowledge from previous tests is known as sequential experimentation and it is a highly recommended practice.

Several outcomes can occur after executing a fractional factorial design. The simplest outcome is that the fractional factorial is resolution V so that the design does not contain any aliasing involving main effects or two-factor interactions. In this case, all conclusions can be made based on this first round of testing, and follow-on testing will simply be for confirmation of the fit of the model.

On the other hand, if the design is resolution III or IV and there is aliasing involving main effects or two-factor interactions, the next steps will differ depending on which factors are determined significant. If all of the statistically significant factors/interactions are not aliased with any others, then this situation can simply be treated as if there was no aliasing. The follow on testing here will once again be used for confirmation runs. Conversely, if any of the statistically significant factors/interactions are aliased with another, then the researcher will not be able to assign the effect on the response to any of the factors. At this point, further investigation must be done to determine which factor or interaction caused the change in the response. For example, if the effect AB is found to be significant, but it is aliased with the effect CD which is found to be not significant (which may occur in a more general case), then it is still impossible to say that AB caused the change in the response and not CD. One possible remedy for this problem is to use subject matter expertise. If the researcher is very familiar with the system and can determine that AB would be the cause over CD, then they may choose to call AB statistically significant and not CD. This assumption must be stated in conclusions and is adding further risk! Be careful when using this technique as the goal of these tests generally assumes that this knowledge would not be known. Another option is to do follow-on testing in such a way that this aliasing structure no longer exists. This is what is called a fold over on a fractional factorial design.

Fold over

By using sequential experimentation, we can perform a fold over to break any aliasing. Different fold overs accomplish different objectives. If initial analysis indicates that only a few main effects are significant, then follow-on testing may only require one factor to be de-aliased from two-factor interactions that contain that factor. A full fold over is done by reversing the signs of all the factors from the first half fraction and executing the new runs as a second half fraction. This type of fold over is most common for Resolution III designs as it will separate aliasing between main effects and two-factor interactions. Another possible approach is to fold over on a single factor, which is performed by reversing the sign of only one factor. This approach will result in a design where the two-factor interactions involving that factor are no longer aliased with other two-factor interactions. This approach is useful when the original design is a Resolution IV design. Figure 4 illustrates a single factor fold over for a 2^{6-2} design.

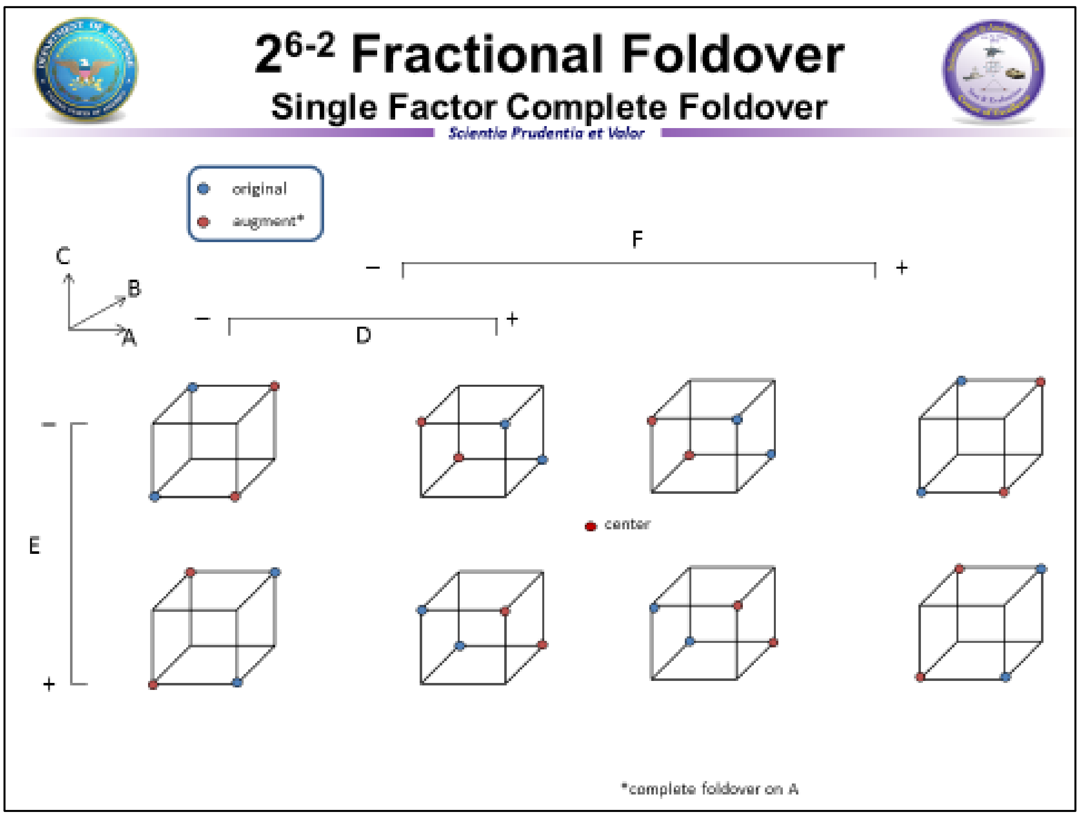


Figure 4: A Fractional Factorial Design + Fold Over example (Blue dots represent the original design and red dots represent the fold over design points)

Conclusion

Full factorial designs are a common starting point when planning a test, but as the number of factors becomes large, the size of the design grows very quickly. Fractional factorial designs are an alternative that can help reduce the number of runs required for screening designs. Many different fractional factorial designs exist, and it is important to note what resolution the design being used is and what aliasing pattern is associated with the design. Several properties, such as the projection property and the sparsity of effects principle, along with sequential testing help make fractional factorial designs such powerful tools. Fractional factorial designs present an opportunity for reduced testing while potentially providing all necessary information, and options to learn about the system in a sequential manner.

References

Montgomery, Douglas C. *Design and Analysis of Experiments.* 9th ed., John Wiley & Sons, Inc., 2017.

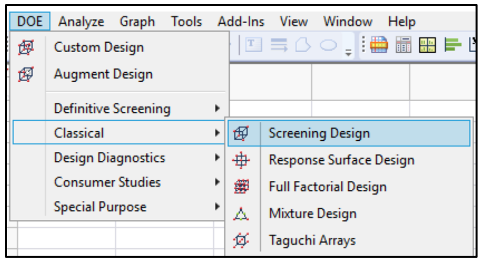
Errore, A., Jones, B., Li, W., & Nachtsheim, C. J. "Benefits and Fast Construction of Efficient Two-Level Foldover Designs." *Technometrics*, vol. 59, no. 1, 2017, pp. 48-57. doi:10.1080/00401706.2015.1124052.

JMP demo

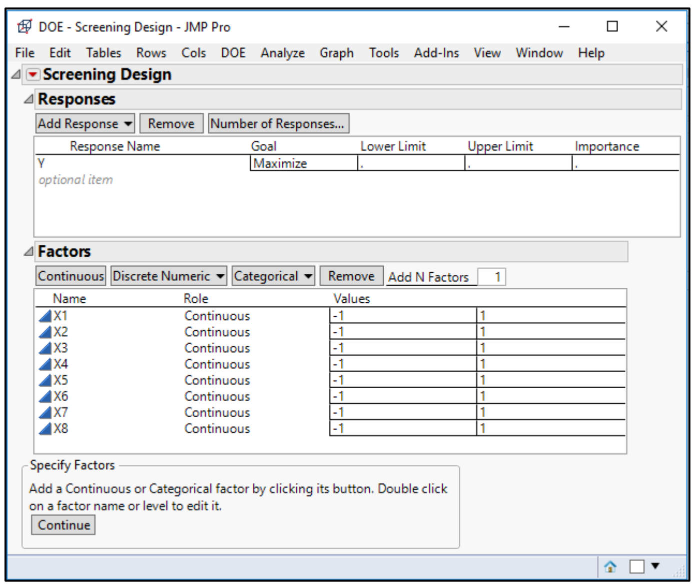
Assume we wish to create a fractional factorial design for a test with 8 continuous factors in JMP 13 (or later version).

1) Open a new data table in JMP.

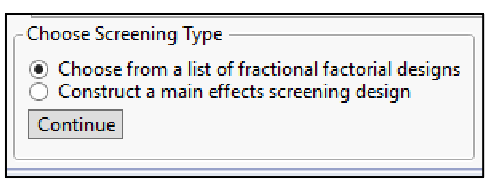
2) Select "DOE -> Classical -> Screening Design".



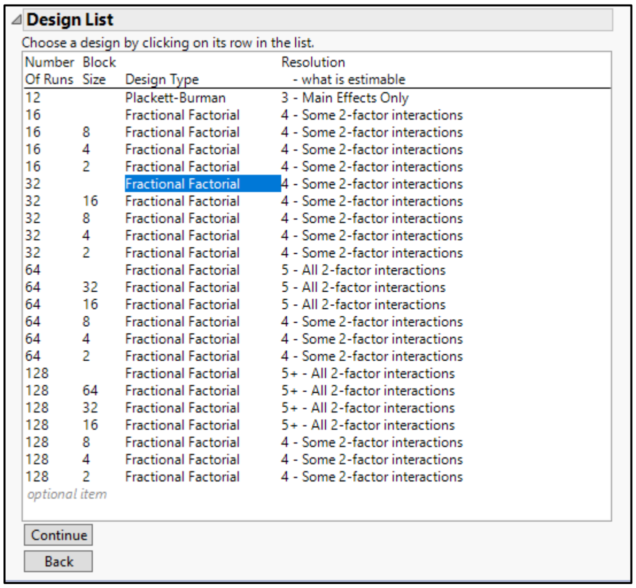
3) Load or enter the factors into the factor section and select "Continue".



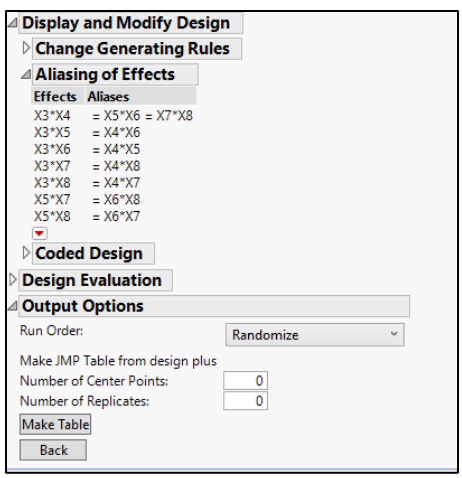
4) Select "Choose from a list of fraction factorial designs" and select "Continue".



5) Select the desired fractional factorial design. Here we will select the 32-run, Resolution IV design with no blocks. This window allows for selecting different sized fractional factorials (such as one-half and one-quarter). There are also designs that include blocking. The resolution is shown in the column on the far right as well as what terms are aliased. Select "Continue".



6) The aliasing structure can be found in this window by selecting the drop down next to "Aliasing of Effects".

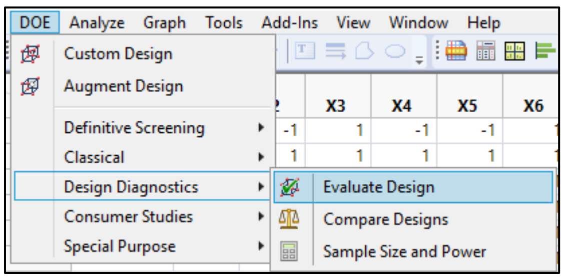


7) Enter the number of Center points desired (typically 3-5). Select “Make Table”.

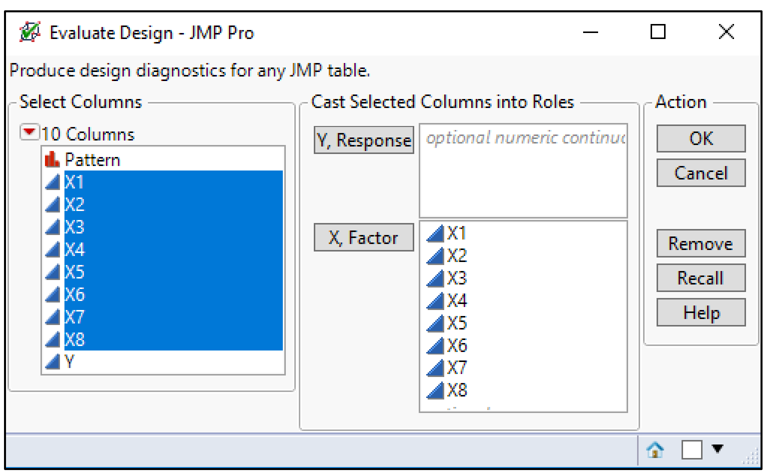
8) The design table is now created and these are the settings for each of the factors in each run. If run order was “Randomize” in the previous step, it can be run in this order. Always randomize!

	Pattern	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	--+--++-	-1	-1	1	-1	-1	1	1	-1	•
2	-+++++--	-1	1	1	1	1	1	-1	-1	•
3	+--+-++-	1	-1	-1	1	-1	1	1	-1	•
4	+-+-+-+-	1	-1	1	-1	1	-1	1	-1	•
5	----++--	-1	-1	-1	-1	1	1	-1	-1	•

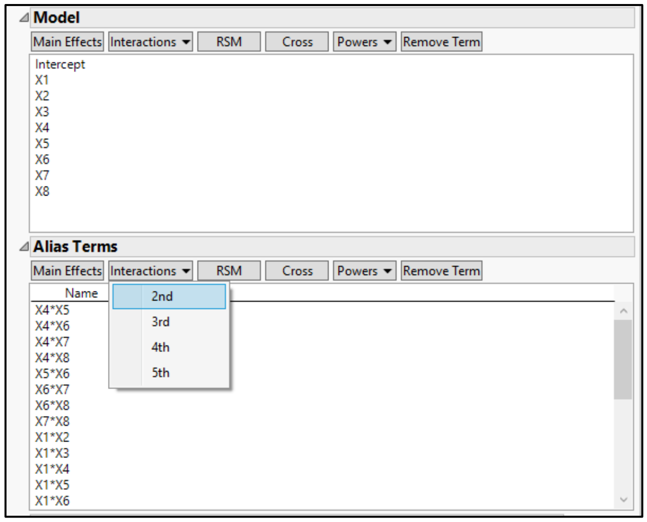
9) In order to view the color map of correlations, select “DOE -> Design Diagnostics -> Evaluate Design”.



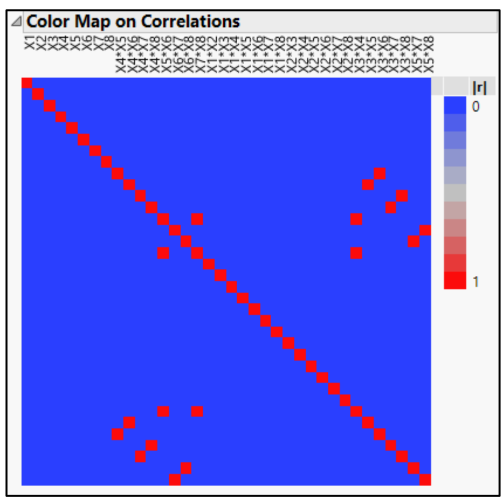
10) Load the factors into the “X, Factor” box and select “OK”.



11) In the model window remove any terms that are not simply main effects or the intercept and in the alias terms window select "Interactions -> 2nd" to include all two-factor interactions.

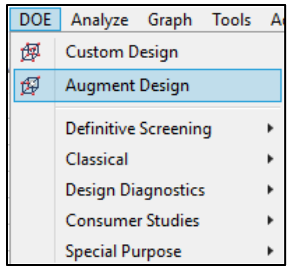


12) Scroll down and select the drop down next to "Color Map on Correlations".

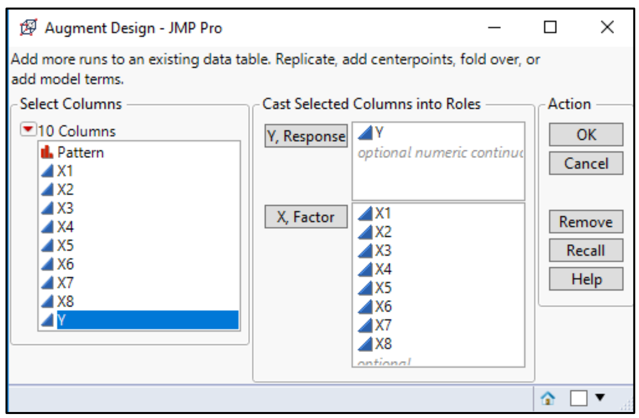


13) Observe the color map for aliasing. This will align with the aliasing seen in the "Aliasing of Effects" window. This also confirms a Resolution IV design as two-factor interactions are aliased with other two-factor interactions, but main effects are not.

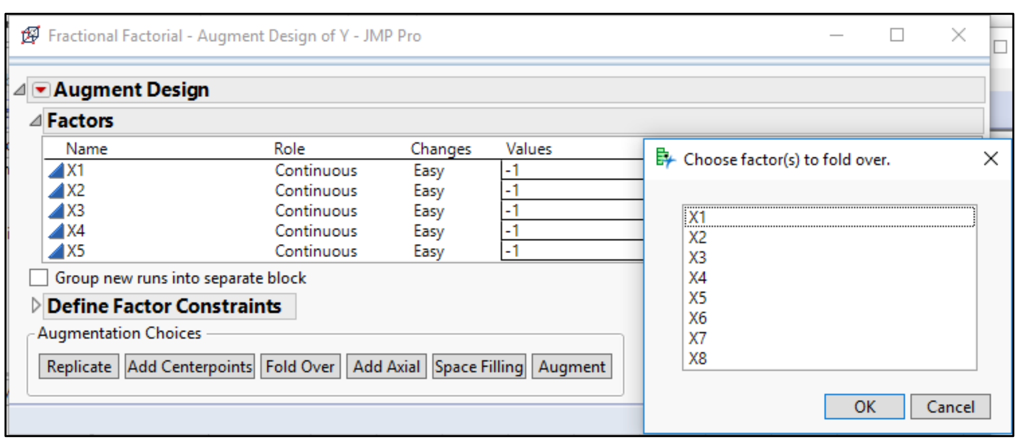
14) If we wish to add runs to the design using a fold over, from the design data table, select "DOE-> Augment Design".



15) Load all of the factors into the "X, Factor" box and the response column into the "Y, Response" box. Select "OK".



16) Select "Fold Over" and choose which factor to fold over. Recall, this is generally done as a second phase of testing if necessary. The factor that would be selected would be the factor that you wish to de-alias.



17) After selecting a factor, select “OK” and the fold over design will be created.

Factor Design

Run	X1	X2	X3	X4	X5	X6	X7	X8
1	-1	-1	1	-1	1	-1	-1	1
2	-1	1	-1	-1	1	1	1	1
3	-1	-1	1	-1	-1	1	1	-1
4	-1	1	1	-1	1	-1	1	-1
5	1	-1	1	-1	-1	1	-1	1

18) Select “Make Table” at the bottom of the window and the new fold over design will be in a JMP table. The first runs will be the previous fractional factorial design created and the additional fold over points will be attached to the end of this designs.

	X1	X2	X3	X4	X5	X6	X7	X8	Y
1	-1	-1	1	-1	1	-1	-1	1	•
2	-1	1	-1	-1	1	1	1	1	•
3	-1	-1	1	-1	-1	1	1	-1	•
4	-1	1	1	-1	1	-1	1	-1	•
5	1	-1	1	-1	-1	1	-1	1	•