More on Planning Experiments to Increase Research Efficiency

Andres Gonzalez

STAT 530

More on Planning Experiments to Increase Research Efficiency

Andres Gonzalez

College of Natural Sciences and Mathematics

California State University – Long Beach

Long Beach, CA 90808

ABSTRACT

This paper clearly demonstrated by example how experimental designs are used to efficiently screen variables in research investigations with little or no mathematical or statistical analysis, and in a minimum number of runs. It will discuss the application of a one-quarter fraction of the 2^k design model to analyze the planning experiments to increase research efficiency for the industrial and engineering chemistry. We want to make sure what factor makes the experiment better and efficient based on the effect of the model. Also, there are useless information flowing around the experiment that is because poorly planned experiments are not likely to contain the information in the first place. To determine the effect of temperature and pressure on per cent conversion in a chemical reaction, a plan that studies temperature at a fixed level of pressure, and then studies pressure at a fixed level of temperature may reveal very little about the real effects of these variables. So, fractional factorial designs are experimental designs consisting of a carefully chosen subset of the experiments runs of a full factorial design which we will be using for this study to require the important information.

INTRODUCTION

There are five factors that will be studied to assess the effect of increase research efficiency. The results can be used in predicting process behavior within the range of variables studied, and in optimizing the process with respect to the variables. Factor A is thought be the condensation of temperature because condensation occurs when the water vapor in the air is cooled, changing from a gas to a liquid which it plays an important role in the water cycle. Factor B is thought to be the amount of material 1 to develop the material that make the technologies we use in our everyday lives better. Factor C is thought to be the solvent volume that depends strongly on the extraction mode used. Factor D is thought to be the condensation time since it can happen at any time of the year and it can occur inside as well as outside of your home. Lastly, Factor E is thought to be the amount of material 2 on yield could increase research efficiency.

METHODS

The experiment was carried out on the eight observations mentioned from the study. We must assume that the factors are independent of each other to base assumption. This was a problem from the book of Design and Analysis of Experiments by Douglas C. Montgomery obtained by the study from an article in *Industrial and Engineering Chemistry* ("More on Planning Experiments to Increase Research Efficiency." 1970, pp. 60-65).

STATISTICAL METHODOLOGY

Design

The 2^{5-2} fractional factorial design is a one-quarter fraction design with two generator word (or the defining relation) is required. The general 2^{k-p} fractional factorial design must select p independent generators meaning that this design will have 2 generators since p=2, and select the first k-p factors to form a full factorial design. On the other hand, a 2^k design with four factors/variables requires 16 experiments for the full replication. Therefore, quarter of the 16 experiments will results in only four experiments producing only three degrees of freedom. Therefore, a quarter fraction design is applicable for five or more factors/variables, so this design have a completed 8 runs rather than 16 since $2^{5-2} = 8$.

Design Construction

To effectively construct a 2^{5-2} fractional factorial design we must know that 2^k refers to designs with k factors where each factor has just two levels as high and low, or +1 and -1. Except for column I, every column has an equal number of + and – signs, and the sum of the product signs in any two columns is zero. Therefore, the product of any two columns yields a column in the table, and each effect has a single degree of freedom.

In our study k = 5, so $2^5 = 32$. Now, we will have two generators of p = 2 since we only have 8 runs with one replication. Therefore, this is a one-quarter fraction 2^{k-2} design. Choosing a generator as I = ACE and I = BDE gives two defining relations as I = ACE = BDE then induce another defining relation as I = ACE × BDE = ABCD. So, the complete defining relation is as I = ACE = BDE = ABCD. Also, based on the word length pattern, we have five factors of resolution III design. Therefore, this a 2_{III}^{5-2} design. Here is the following alias structure:

Alias Structure:
Defining Relation: *I* = *ACE* = *BDE* = *ABCD*
A = *CE* = *ABDE* = *BCD*
B = *ABCE* = *DE* = *ACD*
C = *AE* = *BCDE* = *ABD*
D = *ACDE* = *BE* = *ABC*
E = *AC* = *BD* = *ABCDE*
AB = *BCE* = *ADE* = *CD*
AD = *CDE* = *ABE* = *BC*

Design Layout

To construct the 2^{5-2} fraction design, the first step is to start with a 2^3 full factorial design, and then construct a 2^{5-2} design with I = ACE and I = BDE. So, we get $D = D \times BDE$ = BE and $E = E \times ACE = AC$ same as having it above in the alias structure.

A	B	C	$D=BE$	$E=AC$	
-	-	-	-	+	e
+	-	-	+	-	ad
-	+	-	+	+	bde
+	+	-	-	-	ab
-	-	+	+	-	cd
+	-	+	-	+	ace
-	+	+	-	-	bc
+	+	+	+	+	$abcde$

Use the combination of signs in those factor columns in the table of plus-minus signs for the 2^k design to assign treatment combinations/runs for the effect.

Statistical Analysis

1. *Analysis of variance*

Initial Statistical Model:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$$

Where y is the response variable, $x_1, \ldots, x_5$ are the factors, $\beta's$ are unknown parameters, and $\epsilon \sim N(0, \sigma^2)$ is the random error t erm.

RESULTS

Design Layout

Construction of the One-Quarter Fraction Design of Resolution III with defining relations for this design are *I* = *ACE*, *I* = *BDE* and *I* = *ABCD* yields the following design matrix:

One-Half Fractional Factorial Design for 5 Factors Defining Relation: I = ACE, I = BDE and I = ABCD							
Obs	A	B	C	D	E	Run	y
1	-1	-1	-1	-1	1	e	23.2
2	1	-1	-1	1	-1	ad	16.9
3	-1	1	-1	1	1	bde	16.8
4	1	1	-1	-1	-1	ab	15.5
5	-1	-1	1	1	-1	cd	23.8
6	1	-1	1	-1	1	ace	23.4
7	-1	1	1	-1	-1	bc	16.2
8	1	1	1	1	1	abcde	18.1

Note that since the generators are $I = +ACE$ and $I = +BDE$ this is the principal fraction of this design.

Estimation of effects

The estimation of the effects for the full model of this fraction is given below:

$$l_{effect} = \frac{1}{2^{k-p-1}n}(contrast\ of\ totals)$$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A	-1.52500000	1.57420615	-0.97	0.4349
B	-5.17500000	1.57420615	-3.29	0.0814
C	2.27500000	1.57420615	1.45	0.2853
D	-0.67500000	1.57420615	-0.43	0.7098
E	2.27500000	1.57420615	1.45	0.2853

It is evident that some treatment combinations are negligible. We only included the main effects of the model since the interaction effects of *AB* and *AD* are available to use as error.

Analysis of Variance Table

Consider the percentages that each main effect accounts for in the total sum of squares:

$$SS_{effect} = \frac{1}{2^{k-p}n}(contrast\ of\ totals)^2$$

Source	Sum of Squares	% Contributed
A	4.651	5.18%
B	53.561	59.69%
C	10.351	11.53%
D	0.911	1.02%
E	10.351	11.53%
Total	89.739	

The treatment combinations associated with A, B, C, D, and E account for 88.95% of the main effects.

Also, let's recall the factors:

Factor	Description
Y	Planning Experiments to Increase Research
A	Condensation Temperature
B	Amount of Material 1
C	Solvent Volume
D	Condensation Time
E	Amount of material 2 on yield

With the interaction terms of *AB* and *AD* not considered in the model, the design can be projected into a full 2^3 factorial with only the main factors of A, B, C, D, and E. The results are obtained as follows:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	79.82625000	15.96525000	3.22	0.2537
Error	2	9.91250000	4.95625000		
Corrected Total	7	89.73875000			

R-Square	Coeff Var	Root MSE	y Mean
0.889540	11.57252	2.226264	19.23750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	4.65125000	4.65125000	0.94	0.4349
B	1	53.56125000	53.56125000	10.81	0.0814
C	1	10.35125000	10.35125000	2.09	0.2853
D	1	0.91125000	0.91125000	0.18	0.7098
E	1	10.35125000	10.35125000	2.09	0.2853

From the ANOVA, we can see that at the 5% significance level that the model is not significant relative to the noise level since the F_0 of each independent factor are all less than the critical value of F test. Therefore, we fail to reject the null hypothesis and conclude that the model is insignificant. However, it might be significant at the 10% significance level.

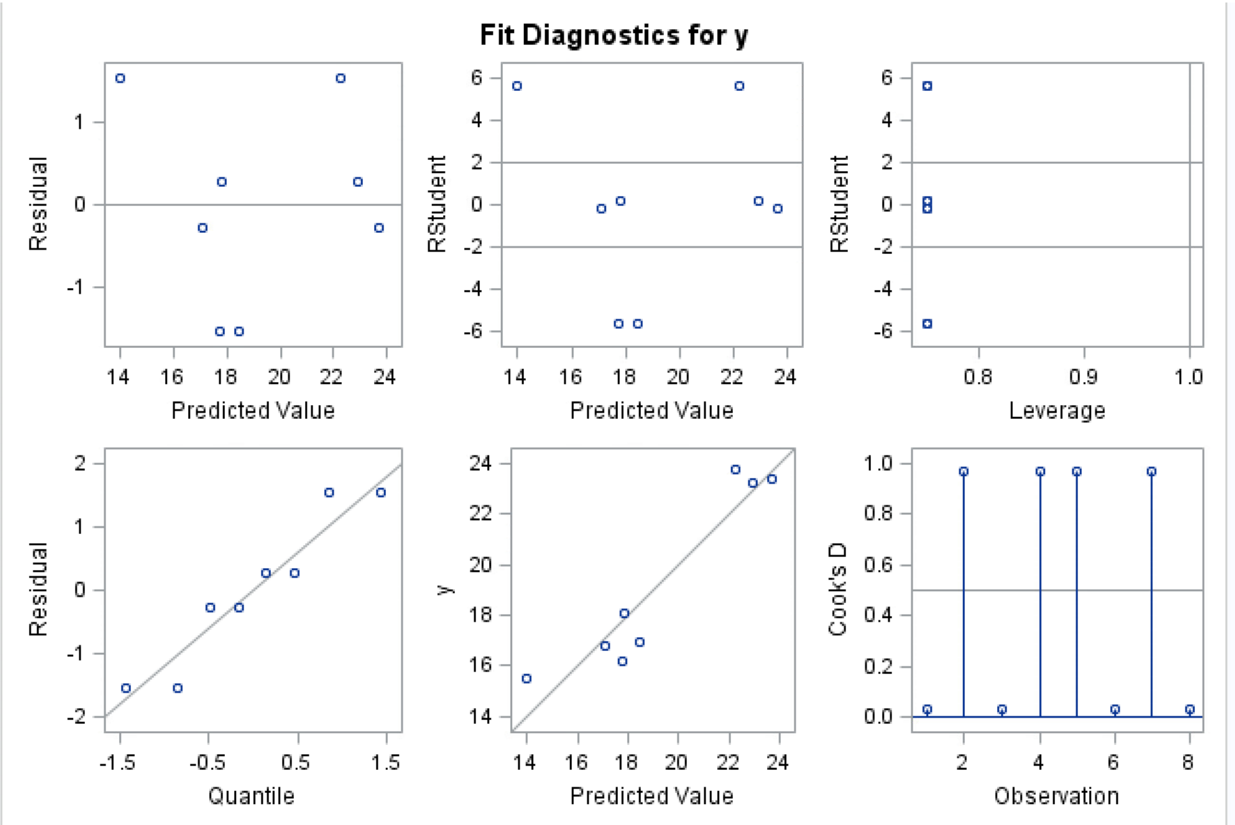
Based on the analysis, the model obtained that can be used to plan experiments to increase research efficiency is as follows:

$$\hat{y} = 19.238 + 4.651x_1 + 53.561x_2 + 10.351x_3 + 0.911x_4 + 10.351x_5$$

The analysis of variance table is shown above. Also, in the previous steps shows that *AB* and *AD* are aliased with other factors. So, if all two-factor and three factor interactions are negligible, then *AB* and *AD* could be pooled an estimate of error.

Plot the Residuals versus the Fitted Values

The results of the plots are obtained as follows:



There appears to be no outliers and is normally distributed since it shows no pattern. Therefore, the residual plot is satisfied and does not violate the assumptions.

Conclusion

The purpose of this experiment is to decide what is more important on planning experiments to increase research efficiency based on the article *Industrial and Engineering Chemistry*. The statistical approach to solve this problem is to use a 2^{5-2} design to investigate the effect of condensation temperature, amount of material 1, solvent volume, condensation time, and amount of material 2 on yield. Based on our results, we conclude that the model is insignificant at the 5% significance level since the F_0 of all independent factors are less than the critical value of F test. Therefore, we fail to reject the null hypothesis for all tests. So, this means that neither of all this factors will not increase the research efficiency of the experiment, and instead a different effect can be included in the study that might be possible to increase the research efficiency, but in that case that would be an another different study to experiment.