Design of Experiments (DOE)

A Brief Overview Paper for
USAF Program Managers

Table of Contents

John C. Sparks
Wright-Patterson Air Force Base
Dayton, Ohio 45433

1) Historical Introduction

The year was 1955. World War II was rapidly fading into the past as America grappled with the new challenges brought by the emerging Iron Curtain. I as a baby boomer was living the idyllic life of a fifties 4th grader straining to catch a TV glimpse of an emerging Elvis—complete with Teddy Bears—in brilliant BW. Simultaneously, we red-blooded American boys were laughing at the "cheap Japanese junk" being passed out as prizes at local elementary schools during the annual fall galas. Little did any of us young machos know that Japan was ready to embark on an industrial revolution, one that would severely challenge American economic dominance—even more so than the Soviets were challenging America's military dominance.

W. Edwards Deming, the acknowledged "Godfather" of the American quality movement, did not always enjoy the prominence that he had during his later years. In fact, Deming was somewhat frustrated with the industrial deaf ears inhabiting America right after World War II, who saw themselves as winners…and kings of the manufacturing world. Made in America was to be made right—no questions asked! So, what do you do with an opinionated genesis? Send him to Japan courtesy of the U.S. government. Perhaps the cantankerous Deming could help rebuild Japan's bombed-out industrial base. While in Japan, Deming befriended Genichi Taguchi, a young Japanese electrical engineer eager to help in the rebuilding effort. Having found a dedicated disciple, Deming quickly trained the bright Taguchi in the twin quality pillars of Statistical Process Control and Factorial Design of Experiments (the subject of this workbook). Taguchi immediately improved the academic presentation of these methods making them readily understandable by other engineers in the struggling Japanese economy. The first big industrial test of Design of Experiments was soon to come.

INA Tile, a Japanese manufacturer of ordinary bathroom tile, had built a brand new kiln at a cost of US $500,000.00. Alas, the poor device had such an uneven heating pattern that all fired tiles broke and crumbled without fail before they could be shipped to market. Talk about being caught between a tile and a hard place! INA was on the verge of bankruptcy with no where to go! Enter the consultant Taguchi who immediately changed the problem from one of rebuilding the kiln to one of making an improved slurry mixture (the tile paste) that could withstand a heavy dose of uneven heating and still transform itself into perfectly good bathroom tile. In today's language, we would call Taguchi's problem change a "paradigm shift" since Taguchi refocused the thinking of INA plant officials to adjusting the slurry, instead of the economically unfeasible alternative of tearing down and rebuilding the kiln.

Pressing onwards towards solution, Taguchi identified eight active ingredients used in the slurry mixture. Each ingredient could be added to the mixture like salt or pepper, in various strengths. Using existing manufacturing records, Taguchi choose two representative strengths for each ingredient that he felt would keep the fired tiles within overall performance specifications. The goal was to find the optimal combination of all eight strength levels (one level per ingredient) leading to durable bathroom tile. "Wait a minute", INA officials complained! "There isn't enough money to carry out your test program. You'll need 2^8 trials or 256 batches of slurry in order to test all possible slurry combinations. We simply can't afford this!" Wrong! Oh so Wrong!

Taguchi was able to conduct a successful test program using a fractional factorial design consisting of 16 individual trials (and a few extra trials thrown in for good measure). INA was saved, and all of Japan rejoiced in the year of 1955. Remember Elvis?

America remained asleep to the powerful new methods of quality until the early 1970s when Japanese manufactured goods imported to this country were suddenly found superior to their American counterparts. The reason was a man named Deming who, two decades before, exported his statistical passions to Japan. The Japanese listened and Made in Japan took on a whole new meaning: quality products reasonably priced and built to last! So what about the present? The experimental design techniques presented in this paper are used around the world wherever quality goods are made. America is definitely no exception; and, in fact, is now considered the quality leader. As an ironic postscript, American industry enticed Genichi Taguchi himself to move to Detroit in the 70s, in order to help revitalize the automobile sector. There, Taguchi founded a company called the American Supplier Institute now run by his son, Shen. In a sense, the original Deming techniques, after undergoing initial testing and refinement in Japan, have returned to America as a fully-matured methodology. As Paul Harvey would say, “Now you know the rest of the story.”

2) Factor-Only Fractional Factorials

2.1 Overview Statement

Fractional factorial experiments are experimental test programs that allow an engineer or technician to obtain statistically valid data using only a small fraction of the available test combinations.

2.2 The Corn-Growing Example

To see how fractional factorial experiments work, let’s consider a corn-growing experiment where the objective is to maximize the yield of corn in bushels per acre. The following are considered important to corn growth: brand of fertilizer, hybrid type, and daily-water amount. These are the *inputs* for the experiment, which is then conducted in order to observe the *measurable* response or output: bushels per acre. Other names for the various *inputs* are *factors* or *stimuli*. Factors are true variables when they are allowed to assume two or more values, called factor levels, during the course of experimentation. Assume that each of our factors in the corn growing experiment has precisely two *levels*, i.e. two different fertilizers, two different hybrids, and two different daily-water amounts. The three factors and their associated levels are shown in **Table 1**.

FACTOR	LEVEL **-1**	LEVEL **1**
Fertilizer	*Top Stalk*	*Fat Ear*
Daily Water	*½ Inch*	*1 Inch*
Hybrid	*Super A*	*Super X*

Table 1: Factor-level Codes

Now think of the corn-growing experiment as a simple manufacturing process whose product is yield.

Prior to making a "manufacturing run", we select one fertilizer *(from two possible choices)*, one water amount, and one hybrid as inputs to the production process. We then turn on the machine—i.e. conduct the experiment—and wait for the measurable product. Each of the factors is thought of as a dial on the side of the machine where dial positions represent the various factor levels. For two-level experiments such as ours, the factor levels are coded **-1** and **1**. The choice is quite arbitrary, but **-1** is usually reserved for the factor level that we think produces the *least* desirable result: in our case, the more modest yield. We will assign the two codes, **-1** and **1**, as shown in the column titles for **Table 1**.

Figure 1 is the corn-growing experiment viewed as a manufacturing process. When the machine is turned on for a given set of three inputs, it runs and hums—*the little gears*—until product is made. We can then reset the three dials at different levels and run the machine again, each time producing a product with, perhaps, a different measurable output.

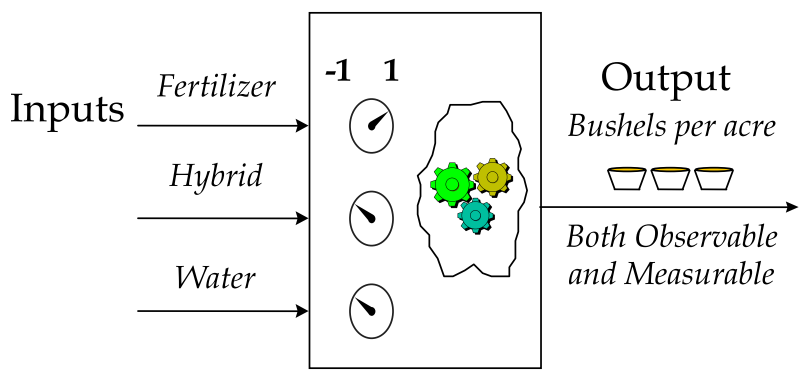

Figure 1: Corn Growing Experiment as a Manufacturing Process

How do we manipulate the dials in order to produce the best possible results? Since the machine has three dials, each having two possible positions, there will be $2^3 = 8$ total possible dial combinations as shown in **Table 2**.

DIAL COM	FERTILIZER	WATER	CORN
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

Table 2: The Eight Possible Dial Combinations

One choice is to simply run all eight combinations *(called full-factorial experimentation)*, observe the eight measurable outputs, and pick the winner. But suppose we had more than three factors, say fourteen? Is this feasible? **Figure 2** depicts a fourteen-dial process.

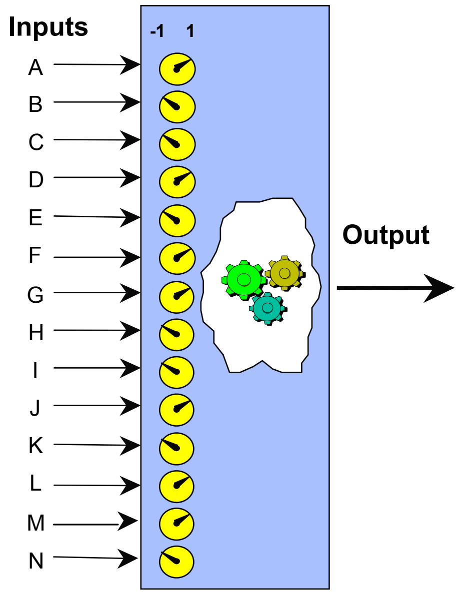

Figure 2: Fourteen-Dial Process

Fourteen was the number of factors that the USAF examined in a major test program, called the Halon Replacement Program (see **USAF Spotlight Box** at the end of this Section), which was conducted to find a new aircraft-fire extinguishing agent. To find the total number of dial combinations for fourteen two-level factors, one computes the expression

$$2^{14} = 16{,}384$$

Can you imagine a table like **Table 2** with $16{,}384$ rows? We definitely have a problem! Who has the time, or money, to run such a massive experiment? Usually we are limited by time and money constraints to a handful of combinations—perhaps fifty—and we better pick the right fifty in order to accomplish the needed job. So, how do we do pick by experience, by guessing, by convenience?

Ronald A. Fisher, a British agricultural statistician, studied the dial problem early in the last century and developed several statistical techniques to deal with this issue. These techniques were further refined through the years and were given the collective name of Design of Experiments (DOE), of which, fractional factorial experimentation is a major component. Dr. William Edwards Deming and Dr. George Box were early proponents of the newly developed DOE technique in the United States. George Box studied under Ronald Fisher, and, in fact, married Fisher's daughter. Dr. Box and his bride returned to the United States and the University of Wisconsin to engage in a lifetime of teaching and writing.

A census bureau statistician, Deming was a recognized expert in sampling and experimental design prior to Word War II. After World War II, Dr. Deming took his expertise (including DOE) to Japan in order to help the Japanese people rebuild their war-torn economy. By 1960, Deming was a household name in Japan, and Japanese quality was rising at a meteoric rate. During his stay in Japan, Dr. Deming mentored Dr. Genichi Taguchi, a young Japanese electrical engineer. Dr. Taguchi worked hard to establish an understandable, systematic DOE methodology that would allow utilization of the powerful fractional factorial techniques by rank-and-file engineers. "Taguchi methods" refers to the fractional factorial system established and formalized by Dr. Taguchi. Taguchi's system makes clever and very efficient use of visual charts, diagrams, and tables. It is a system very much preferred by engineer practitioners.

Continuing our discussion of the dial problem, let's use a mathematical metaphor. When you studied high-school geometry, you were first introduced to "axioms" which are the building blocks of the entire geometric system. All theorems, no matter the complexity, were proved using these "axiomatic" building blocks. The knowledge contained in these building blocks allowed us to derive the totality of the knowledge within the system. *Enter DOE and fractional factorial experimentation.* By utilizing special mathematical arrays *(matrices*) having unique properties, the DOE methodology (as developed by Fisher, Box, Deming, Taguchi, etc.) allows the investigator to systematically select an "axiomatic" subset (or fraction!) of dial combinations that will form the backbone of the test program. In the case of the Halon Replacement Program, 32 dial combinations were needed, due to the many more factors, to build this "axiomatic" subset. The data generated from the "axiomatic" subset was then used to mathematically construct results for the remaining $16{,}352$ runs! Notice that we are getting the needed results from a very small fraction ($1/512$) of the $16{,}352$ possibilities. A series of special mathematical arrays called orthogonal *(orthogonal is a fancy name for perpendicular)* arrays are the key DOE tools that allow an investigator to pick precisely the right fraction which defines the "axiomatic" subset.

Table 3 is an example of one of these arrays having four rows and four columns with designation L4 *(L is the last letter in the word orthogonal)*. Other sizes available for experiments that have two-level factors are L2, L8, L16, L32, L64, and L128: notice that two-level orthogonal arrays progress by powers of 2. *For the curious, the series for three-level factors is L9, L27, L81, and L243.* The L4 shown on the next page is used for any experiment consisting of four separate "axiomatic" trials where each row is a code for running one of the trials; each column, a code for analyzing the resulting data. We are going to use the L4 array to both conduct and analyze our corn-growing experiment.

1	-1	-1	-1
1	-1	1	1
1	1	-1	1
1	1	1	-1

Table 3: L4 Orthogonal Array

To start, assign a two-level factor to each column in this array, except the first column of all bolded **1**s. As shown later, the first column is reserved for calculating the output average for all the axiomatic trials. Since we have three two-level factors, we can make column assignments as shown in **Table 4**.

AVERAGE	FERT	WATER	CORN
1	-1	-1	-1
1	-1	1	1
1	1	-1	1
1	1	1	-1

Table 4: Column Assignments

The next step is to add a fifth column called bushels *(the output variable)* and replace the -1's and 1's in the factor columns with the actual factor levels. Using the factor-level codes from **Table 1**, **Table 5** shows the result known as the execution array.

AVERAGE	FERT	WATER	CORN	*BUSHELS*
1	*Top Stalk*	*½ inch*	*Super A*	
1	*Top Stalk*	*1 inch*	*Super X*	
1	*Fat Ear*	*½ inch*	*Super X*	
1	*Fat Ear*	*1 inch*	*Super A*	

Table 5: Execution array

Each row in the execution array tells the investigator how to set one of the four axiomatic-dial combinations in the corn-growing experiment. For example, row 3 tells the investigator to conduct a test using *Fat Ear* fertilizer, a daily water amount of *½ inch*, and the *Super X* hybrid throughout a *standardized growth period*. At the end of this trial, the results for row 3 are recorded in the *bushels* column. This process is repeated for rows 1, 2, and 4. We do not have to conduct the corn-growing experiment using the row sequence 1,2,3,4. In fact, this experiment would probably run in parallel using four adjacent growing fields. **Table 6** shows the completed execution array.

AVERAGE	FERT	WATER	CORN	*BUSHELS*
1	*Top Stalk*	*½ inch*	*Super A*	120
1	*Top Stalk*	*1 inch*	*Super X*	109
1	*Fat Ear*	*½ inch*	*Super X*	128
1	*Fat Ear*	*1 inch*	*Super A*	46

Table 6: Completed Execution Array

The entries in the *Bushels* column will now allow the investigator to develop a linear equation that expresses bushels in terms of a simple linear combination of fertilizer (F), Water (W) and Corn (C):

$$Bushels = c_0 + c_1 F + c_2 W + c_3 C$$

To generate the four coefficients c_0, c_1, c_2, c_3 go back and replace the named factor levels in **Table 6** by their **-1** and **1** equivalents from **Table 4** in order to obtain **Table 7**, which is called the analysis array.

AVERAGE	FERT	WATER	CORN	*BUSHELS*
1	-1	-1	-1	120
1	-1	1	1	109
1	1	-1	1	128
1	1	1	-1	46

Table 7: Analysis Array

Each coefficient is generated using the column to which the associated factor is assigned and all four of the data values: 120, 109, 128, and 46. Since c_0 is not associated with a variable, it is generated using the AVERAGE column and the *BUSHELS* column in the following fashion

$$c_0 = \frac{1x120 + 1x109 + 1x128 + 1x46}{4} = 100.75$$

The divisor of 4 corresponds to the 4 data points. Note that c_0 is simply the average yield from all four tests. To obtain the coefficient c_1, we perform a similar calculation using the FERT*ilizer* column and *BUSHELS* column:

$$c_1 = \frac{(-1)x120 + (-1)x109 + 1x128 + 1x46}{4} = -13.75$$

Generating c_2 and c_3 in like fashion, the completed equation becomes:

$$Bushels = 100.75 - 13.75F - 23.25W + 17.75C$$

The above equation will reproduce any data point in the original data set. For example, Row 3 produced the value 128 when the experiment was run at the levels **1**, **-1**, and **1** *(Fat Ear, ½ inch, Super X)*. If we substitute **1**, **-1**, and **1** into the above equation, the result becomes:

$$Bushels = 100.75 - 13.75x(1) - 23.25x(-1) + 17.75x(1) = 128$$

Notice the last result matches the original data point, as it should!

Going back to the original dial problem, how do we choose the winner? How do we choose the right level of fertilizer, water amount, and hybrid in order to maximize the yield in bushels? The equation itself is the key to the answer. If our objective is to make $Bushels$ as large as possible, the way do this is to set $F = -1$ *(Top Stalk)*, $W = -1$ *(½ inch*), and $C = 1$ *(Super A)*. Substituting the three values **-1**, **-1**, and **1**, we obtain:

$$Bushels = 100.75 - 13.75x(-1) - 23.25x(-1) + 17.75x(1) = 155.5$$

Notice that the maximum yield in bushels occurs at a dial combination not in the original set of four. The final step is to perform a confirmation run into order to “prove” the value generated by the equation. Since three two-level factors have eight possible dial combinations, and we ran four axiomatic-dial combinations to obtain our equation, the experimental design is called a ½ fractional factorial. This was exactly the same process (in miniature) as the process used in the Halon Replacement Program. In the Halon Replacement Program, since a L32 (32 axiomatic-dial combinations) was used to examine the 16,384 possible dial combinations, the associated experimental design was called a 1/512 fractional factorial.

What are some methods that we can use to visualize the data? One of the easiest is a bar chart, **Figure 4** below, showing relative factor strength as measured by the absolute value of the factor coefficient.

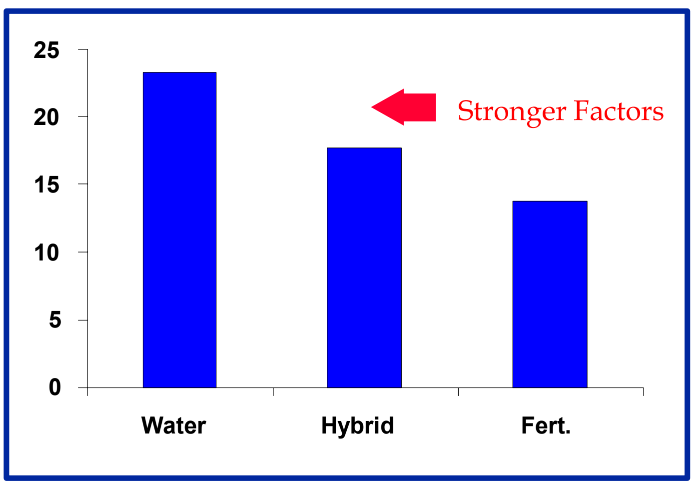

Figure 4: Relative Factor Strength

Other methods of visualizations are also available such as an Effects Diagram and Cube Plot(s), but **Figure 4** is about as simple as you get and it aligns well in structure with existing Pareto displays.

2.3 DOE Summary Points

In this section, we have used DOE and a fractional-factorial design to solve a problem in agriculture. Additionally, we have integrated the mathematics of algebra with some very practical biology. As you can see, the fractional factorial technique was used to save the investigator some very valuable time and money. In this case, the corn-growing experiment was conducted at one-half the cost of an equivalent “full-factorial” test program. Can you imagine the money saved in the USAF Halon Replacement Program!

The success of fractional factorial experimentation depends, in part, on the skill of the investigator in identifying the important factors that have potential for driving the process under investigation. Of particular importance is the early identification of potentially potent combinations of two variables that might influence the "manufacturing process". These two-factor interactions are then assigned to columns just like factors, and the experiment, though now more complicated, proceeds in a generally like fashion.

2.4 USAF Success Story

The Air Force Research Laboratory (AFRL), headquartered at Wright-Patterson Air Force Base in Dayton, Ohio, is home to more than 6000 scientists and engineers responsible for maintaining America's technological pre-eminence in the air! New technology needs to be thoroughly tested within budgetary constraints before it is released into the Air Force inventory. AFRL uses Design of Experiments (DOE) and the associated fractional-factorial methodologies as a way to achieve both objectives.

For example, in the mid 90s, the Air Force embarked on a program to replace the then-used Halon fire-extinguishing agent with an alternative agent, one that would not harm the earth's ozone layer. The selected replacement agent had to be operational before the year 2000 with an extinguishing capability matching or exceeding Halon. Several million dollars were allocated to a program with the goal of finding and thoroughly testing an alternative fire-extinguishing agent. Initial brainstorming revealed several dozen factors that influence the propensity of an aircraft to catch on fire, and the associated ability of an extinguishing agent to put the fire out. The factors were then rank-ordered by importance; and, from the rank-ordered factors, the first 14 were selected for actual testing. Each factor, such as airflow velocity, had a least two different values, called levels. If all possible factor-level combinations had actually been tested, $2^{14} = 16,384$ individual tests would have been needed. Budgetary and time constraints definitely did not allow for this "full-factorial" and extremely costly approach. Fortunately, AFRL scientists and engineers were knowledgeable in fractional-factorial experimental design methods. They were able to construct a highly efficient, statistically valid test program consisting of 32 individual tests. This program, once conducted, would allow for the precise extraction of the needed information at a cost that was $1/512$ the cost of the corresponding full factorial! In retrospect, the **Halon Replacement Program** proved to be a major success, generating a technically valid product delivered well within the time and budgetary constraints. Design of Experiments and associated fractional-factorial methodologies helped made it so.

Major AFRL test programs are not the only opportunities available for utilization of DOE. Throughout the years, both big and small AFRL test programs have been successfully conducted using the methods introduced in this paper. DOE has been the cornerstone of many successful test programs advancing the following aircraft technologies—sensors, aerodynamics, high-strength micro-materials, large structures, propulsion, power systems, human subsystems, pilot readiness and training, and weapon system delivery and effectiveness—to name a few.

3) The Analysis of Variance (ANOVA)

3.1 Overview Statement

ANOVA is a statistical-based error methodology (actually based on the Pythagorean Theorem!) that one can use in conjunction with DOE and fractional-factorial experimental designs. ANOVA can only be used when there is a valid way to ascertain random error from the experimental results. In the DOE toolkit, there are several different powerful methods that will allow a researcher to make an estimate of this error. In this brief introduction, we will look at only one of these methods, called replication.

Once an error estimate is made, it is given the name "noise". Likewise, associated estimates of factor and interaction (ALT 6) strengths are called "signals". ANOVA will allow us to statistically examine the magnitude of signal-to-noise ratios in order to see if factor/interaction strengths are truly significant.

3.2 Continuing the Corn-Growing Example

To see how ANOVA works, reconsider the corn-growing experiment where each of the four trials has been replicated (i.e. repeated for two rounds). **Table 8** below shows the final result.

AVERAGE	FERT	WATER	CORN	*BUSHELS ROUND 1*	*BUSHELS ROUND 2*
1	*Top Stalk*	*½ inch*	*Super A*	120	114
1	*Top Stalk*	*1 inch*	*Super X*	109	111
1	*Fat Ear*	*½ inch*	*Super X*	128	123
1	*Fat Ear*	*1 inch*	*Super A*	46	53

Table 8: Completed Execution Array with Two Replications

As before, the investigator can develop a linear equation from the associated analysis array that again expresses bushels in terms of a simple linear combination of Fertilizer (F), Water (W) and Corn (C):

$$Bushels = c_0 + c_1 F + c_2 W + c_3 C$$

Table 9 is the associated analysis array for our replicated experiment.

AVERAGE	FERT	WATER	CORN	*BUSHELS ROUND 1*	*BUSHELS ROUND 2*
1	-1	-1	-1	120	114
1	-1	1	1	109	111
1	1	-1	1	128	123
1	1	1	-1	46	53

Table 9: Analysis Array with Replications

As explained in Section 2, each of the four coefficients c_0, c_1, c_2, c_3 is generated using the column to which the associated factor is assigned. But now, all eight data values 120, 109, 109, 46 and 114, 111, 123, 53 are used. To generate the constant coefficient c_0, use the AVERAGE column and the two *BUSHELS* columns as shown in the following fashion:

$$c_0 = \frac{1x[120+114]+1x[109+111]+1x[128+123]+1x[46+53]}{8} = 100.5$$

The divisor of 8 corresponds to our 8 data points. Note that again c_0 is simply the average yield from all eight tests. The remaining three coefficients are done in analogous fashion. To obtain c_1 c1, use the FERT column and the two *BUSHELS* columns as shown:

$$c_1 = \frac{(-1)x[120+114]+(-1)x[109+111]+1x[128+123]+1x[46+53]}{8} = -13.0$$

Generating c_2 and c_3 in like fashion, the completed equation becomes:

$$Bushels = 100.5 - 13.0F - 20.75W + 17.25C$$

The above equation will now reproduce *the average* for any trial in the original data set. For example, Row 3 produced (as the result of two replications) an average value of 125.5. When the levels **1**, **-1**, and **1** *(Fat Ear, ½ inch, Super X)* are substituted into the above equation, the result becomes:

$$Bushels = 100.5 - 13.0x(1) - 20.75x(-1)W + 17.25x(1) = 125.5$$

The calculated value matches the average of the two Row 3 data values. To finish this analysis, one should construct some sort of visual display such as a Bar Chart, Effects Diagram, or Cube Plot.

Going back to the original problem, how do we know if the three factor coefficients are large enough to represent a true signal? *Enter ANOVA and error analysis*. To start, look at the two values associated with Row 1: 120 bushels and 114 bushels. Why are these two values different? The difference is due to all of the untested factors that come into play when we ran the corn-growing experiment. These untested and hidden factors are haphazardly mixed in an unknown fashion and constitute the random error in the experiment. One estimate of this random error would simply be the range of $120-114=6$ associated with the two data points. ANOVA refines this initial estimate by squaring the 6 and then dividing by 2 (the number of data values per row) to obtain a noise estimate for Row 1 of 18. Similar calculations for Rows 2, 3, and 4 result in the values 2, 12.5, and 24.5. Each of the four values 18, 2, 12.5, and 24.5 is an independent estimate of the random experimental error and has been generated under a unique set of factor/level combinations.

Notice that these four estimates vary widely. The statistician uses the average of the four estimates, 14.25, to measure the *sum of squares for noise*. Since four ranges were used to build this average, we say this average sum of squares has *four degrees of freedom.*

How do we calculate the signals for each of the three factors and overall average? Simply take each of the four coefficients, square it, and then multiply by 8 (the total number of data points). The resulting number is called the sum of squares to the associated factor effect and is a measure of *signal strength*. For two-level factors, each signal will have *one degree of freedom*. The sum of squares for the overall average coefficient c_0 also has *one degree of freedom*. For any given experiment, the total for degrees of freedom is simply equal to the total number of data points. In our example, the total for degrees of freedom is eight. In a replicated experiment, such as ours, half of the degrees of freedom are assigned to error or noise. The remaining degrees of freedom are divided evenly amongst the overall average and factor(s) sum of squares. More complicated schemes for divvying up the total degrees of freedom exist. These partitioning schemes will depend on how many factor levels we have and the statistical nature of our noise estimate.

Table 10 (called an ANOVA table) arranges all of the important signal and noise data from the previous calculations. **In Table 10**, DF is the abbreviation for degrees of freedom.

AVERAGE OR FACTOR	SUM OF SQUARES	DF	SIGNAL DIVIDED BY NOISE	TRIP VALUE AT 99% SIGNIFICANCE
Average	80,802.	1	5670.3	21.2
Water	3444.5	1	237.55	21.2
Corn	2380.5	1	164.17	21.2
Fertilizer	1352.0	1	93.24	21.2
Noise	14.25	4	Skip	Skip

Table 10: ANOVA Table for the Replicated Experiment

In Column 1, the overall average and each experimental factor/interaction are given a row. The noise sum-of-squares estimate is also given a row at the bottom, preferably separated from the potential signal sum-of-squares estimates by an empty row, since the sum of squares for noise is to serve as the common divisor. Sum-of-squares calculations are entered in Column 2. The associated degrees of freedom for each row are entered in Column 3. The Column 4 values are the actual signal-to-noise ratios.

Signal-to-noise ratios deserve more explanation. How big does a signal-to-noise ratio have to be in order to “prove” that the signal is not due to the same random factor effects that created the noise? Sir Ronald Fisher, a British statistician, did extensive research on this topic in the early 1900s and developed a series of tables (called F tables) that allow an experimenter to answer this question.

All F tables are constructed for a given significance level (to be explained) and set up so that entries in the body of the table are a function of the number of degrees of freedom represented in both the numerator (average/factor/interaction) and denominator (noise). **Table 11** below is a condensed Fisher (or F) table, containing trip values corresponding to 90% significance, 95% significance, and 99% significance. **Table 11** is good for signal-to-noise ratios having 1 degree of freedom in the numerator, characteristic of two-level factorial designs.

Denominator Degrees of Freedom	90% Trip Value	95% Trip Value	99% Trip Value
1	39.9	161.4	4052.00
2	8.53	18.51	98.5
3	5.54	10.13	34.1
4	4.54	7.71	21.2
5	4.06	6.61	16.3
6	3.78	5.99	13.8
7	3.59	5.59	12.3
8	3.46	5.31	11.3
9	3.36	5.12	10.6
10	3.28	4.96	10.0
15	3.07	4.54	8.68
20	2.97	4.35	8.10
infinite	2.70	3.84	6.63

Table 11: Fisher Table for Three Significance Levels And One Degree-of-Freedom in the Numerator

Significance levels are simply the probability that sum-of-square size is actually due to the strength of the tested factor, and not to the same random unknown factor effects causing the noise. When we say that a factor is 90% significant, we are 90% certain that the sum of squares exhibited by that factor is due to the strength of that factor, and not to unknown random factor effects. Another way of looking at it is that we only have a 10% chance of being wrong. Associated with each numerator degree of freedom, denominator degree of freedom, and *experimenter-chosen* significance level is a so-called trip value for the signal-to-noise ratio. If the calculated signal-to-noise ratio exceeds the trip value, then the signal is significant. *Trip values get larger as the significance level increases. When degrees of freedom increase, trip values get smaller*. In our example, the trip value for a denominator with 4 degrees of freedom, a numerator with 1 degree of freedom, and a 99% chosen significance level is 21.2. All factors exceed the trip value with plenty of room to spare! Hence, in this experiment, they are all highly significant exceeding 99%. Be aware that this is not always the case. The overall average is usually significant no matter what. Its significance is due to the fact that we actually performed an experiment and got something that we could measure; where, prior to the experiment, we had nothing.

A note for the mathematically curious: both fractional factorial Arrays and DOE work as they do because of basic Pythagorean properties that extend to multidimensional space. These properties enable the minimal orthogonal basis as represented in the orthogonal array to cleanly partition the test data into components, where each component is the amount attributed to just one factor or two-factor interaction. The process is similar to that of resolving a velocity vector into horizontal and vertical components using trigonometry.

4) Vane-Cleaning Example

4.1 Problem Statement

A particular gas-turbine engine vane becomes corroded during service and requires periodic cleaning. Very high pressure water is delivered through a tiny nozzle orifice in order to cleanse the vanes. An experiment is designed and conducted in order to maximize cleansing effectiveness. The response variable—to be minimized—is percent contamination remaining after the cleansing procedure.

4.2 Professional Solution and Analysis

If using a DOE-based experimental test program, contractors will typically include in their reports to the USAF the following types of information: **Table 12**, **Figure 6**, **Table 13**, and **Table 14**. It will be up to you, the USAF project engineer, to ask for additional data, which allows for the pealing back of the onion. Many DOE workshops are currently available to help you fill in the information gaps deliberately shown below if you desire or have a need to do so.

FACTOR	**-1** LEVEL	**1** LEVEL
O: Orifice Size	0.007”	1.0”
S: Standoff Distance	0.5”	1.0”
P: Pressure	20KSI	35KSI
F: Feed Rate	20ipm	30ipm
R: RPM	1500rpm	2000rpm

Table 12: Factors and Levels

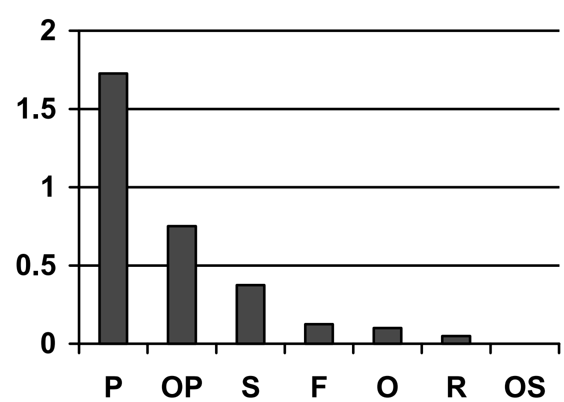

Figure 6: Factor and Interaction Strengths

Note: The OS and OP interactions are thought to be significant. Interactions are particularly potent combinations of two or more factors akin to drug interactions. Interactions are mentioned but not treated in this paper

Grade C	*Fr*	*PF*	*PR*	*SF*	*SR*	*OR*	*SP OF*	<Alias
GM	**O**	**S**	**OS**	**P**	**OP**	**F**	**R**	**%**
1	-1	-1	-1	-1	-1	-1	-1	10.1
1	-1	-1	-1	1	1	1	1	11.9
1	-1	1	1	-1	-1	1	1	9.2
1	-1	1	1	1	1	-1	-1	11.3
1	1	-1	1	-1	1	-1	1	8.9
1	1	-1	1	1	-1	1	-1	13.5
1	1	1	-1	-1	1	1	-1	7.8
1	1	1	-1	1	-1	-1	1	13.1

Table 12: L8 Orthogonal Array for Vane Cleaning Example Showing Confounding Pattern and Design Grade

Factor	**Sum(1)**	**Sum(-1)**	c_i	**SS**
GM	85.8	0.0	10.725	920.205
O	43.3	42.5	.1	.08
S	41.4	44.4	-.375	1.125
OS	42.9	42.9	0	0
P	49.8	36	1.725	23.805
OP	39.9	45.9	-.75	4.5
F	42.4	43.4	-.125	.125
R	43.1	42.7	.05	.02
Totals				*949.86*

Source	**DF**	**SS**	F_{ratio}	**Significance**
GM	1	920.205	19,051.86	>> 99%
P	1	23.805	492.85	99%
OP	1	4.5	93.16	99%
S	1	1.125	23.29	95%
O	1	.08	*must include*	
Error	3	.043	*divisor*	

Table 13: Detailed ANOVA Table

Linear Model and Optimal Setting:

$$C\% = 10.725 + 1.725P - 0.75OP - 0.375S + 0.1O$$

For the optimal setting which minimizes percent contamination, set $P = -1$. Set $OP = -1$, which implies $O = 1$. Finally set $S = 1$ to obtain

$$C\% = 10.725 + 1.725x(-1) - 0.75x(1) - 0.375x(1) + 0.1x(1) .$$

DOE Resources

Books on Statistics

1) Dictionary/Outline of Basic Statistics, Freund and Williams, Dover Publications, 1991
2) Statistics in Plain English, Harvey Brightman, South Western, 1986
3) Statistics for Business and Economics; Anderson, Sweeny, Williams; West, 1994

Books on DOE

4) Statistics for Experimenters: Box, Hunter, and Hunter; Wiley Interscience, 1978
5) Taguchi Techniques for Quality Engineering (2^{nd}), Philip Ross, McGraw Hill, 1996
6) Understanding Industrial Experimentation, Dr. Donald Wheeler, SPC Press, 1988

DOE Software

7) DOE PAC by PQ Systems of Miamisburg, Ohio: call (937) 885-2255
8) Microsoft Excel can be set up to do DOE: AFRL provided diskettes
9) Other Software: Statgraphics, Design Cube, MathCAD