CHAPTER ONE

Introduction

The purpose of this introduction to the Design of Experiments (DOE) is to showcase the power and utility of this statistical tool while teaching the audience how to plan and analyze an experiment. It is also an attempt to dispel the conception that DOE is reserved only for those with advanced mathematics training. It will be demonstrated that DOE is primarily a logic tool that can be easily grasped and applied, requiring only basic math skills. While software would make the calculations more painless and provide greater versatility, it is necessary to understand what the software is doing. To this end, software is not used with this text, but calculators are used instead to insure that the basics are learned. At the conclusion, software applications will be obvious and some options among available packages will be described. This is by no means a complete treatment of the broad field of DOE. The intent is to introduce the basics, persuade the reader of the power of this tool, and then recommend resources for further study. The material covered will still be sufficient to support a high proportion of the experiments one may wish to perform.

The prerequisites of this book are familiarity with the concepts of process stability, basic statistical process control (SPC), and measurement analysis. As in any process improvement activity, it is necessary to recognize that a process is made up of input variables, process variables, and output measures (see Figure 1.1). The intent is always to improve the output measure, which is labeled as the *response.* There is no direct control on the response variable; in the classical cause-and-effect approach, it is the effect.

The *causes* are what dictate the response. To control the response, one must control the causes, which may be input variables and/or process variables involving the five elements shown in Figure 1.1. (These variables or causes will later be referred to as *factors.*)

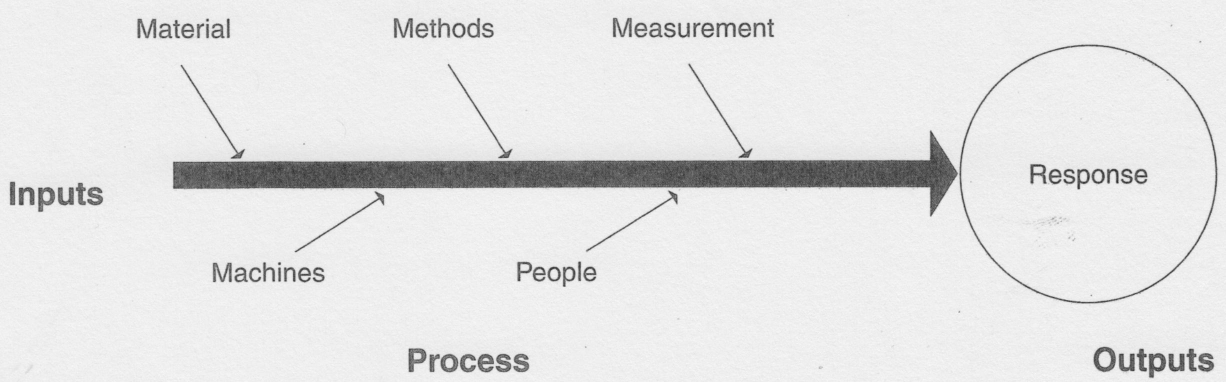


FIGURE 1.1. Cause-and-effect relationship.

For example, there is no control setting in a sales process that allows one to set a sales level. To control sales, one must address those variables that *cause* sales to change, e.g., promotional literature, call frequency, pricing policies, credit policies, personal sales techniques, etc. A process may be very simple, or it may be a complex group of processes.

In concert with this cause-and-effect or systems approach to the process, the concepts of process variation must be understood. Every response demonstrates variation. This variation results from (a) variation in the known input or process variables, (b) variation in the unknown process variables, and/or (c) variation in the measurement of the response variable. The combination of these sources results in the variation of that response. This variation is categorized by the classic SPC tools into two categories: (a) *special-cause variation*—unusual responses compared to previous history; and (b) *inherent variation*—variation that has been demonstrated as typical of that process.

A side note is needed here on terminology. Inherent or typical variation has a variety of labels that are often used interchangeably. In control charting, it is referred to as *common-cause* variation. In control systems, it is called *process noise*. In DOE, it is called *experimental error* or *random variation*. To minimize confusion, it will be referred to in this text as either inherent variation or experimental error.

Control charts are used to identify special-cause variation and, hopefully, to identify the process variables or causes that led to such unusual responses. The presence of special causes within an experiment will create problems in reaching accurate conclusions. For this reason, DOE is more easily performed after the process has been stabilized using SPC tools. The presence of inherent variation also makes it difficult to draw conclusions. (In fact, that is one of the definitions of statistics: decision making in the presence of uncertainty or inherent variation.) If a process variable causes changes in the response that exceed the inherent variation, we state that the change is *significant*.

Inherent variation can also be analyzed to determine if the process will consistently meet a specification. The calculation of process capability is a comparison of the spread of the process with the specifications, resulting in test statistics such as C_p and C_{pk}. Figure 1.2 illustrates the comparison of a process with its upper and lower specification limits.

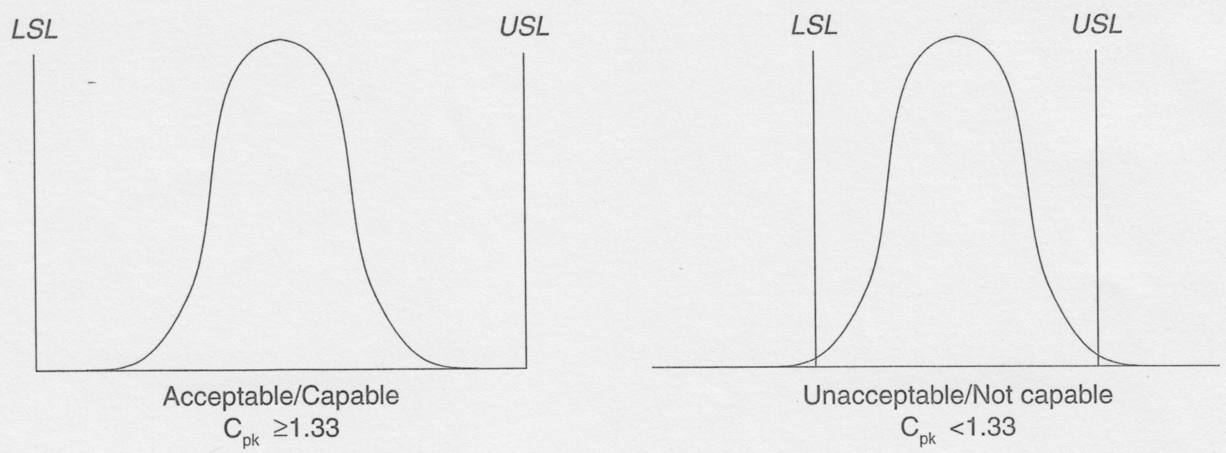


FIGURE 1.2. Process capability.

Design of Experiments is the simultaneous study of several process variables. By combining several variables in one study instead of creating a separate study for each, the amount of testing required will be drastically reduced and greater process understanding will result. This is in direct contrast to the typical *one-factor-at-a-time* approach or *OFAT*, which limits the understanding and wastes data. Additionally, OFAT studies can not be assured of detecting the unique effects of combinations of factors (a condition later to be defined as an *interaction*).

Design of Experiments includes the entire scope of experimentation, including defining the output measure(s) that one desires to improve, the candidate process variables that one will change, procedures for experimentation, actual performance of the experiment, and analysis and interpretation of the results. The objectives of the experimenter in a DOE are

1. to learn how to change a process average in the desired direction
2. to learn how to reduce process variation
3. to learn how to make a process *robust* (i.e., make the response insensitive to uncontrollable changes in the process variables)
4. to learn which variables are important to control and which are not

For ease of instruction, small experiments are presented first, followed by large experiments. In the real world, one would prefer to start with large experiments and progress to smaller ones in order to identify variables that affect the response variables. Terms and definitions are covered as they arise. The terminology used in DOE is often different from the equivalent terms in SPC and is presented to assure that other references are more readable. The initial example is used to define most of the unique terminology and much of the analytical technique.

CHAPTER TWO

Experiments with Two Factors

It is beneficial at this point to review the logic steps to be followed in planning and implementing a DOE, even if some of the terminology has not been defined (this is repeated after the first exercise):

1. Define the process to be studied.
2. Determine the response(s).
3. Determine the measurement precision and accuracy.
4. Generate candidate factors.
5. Determine the levels for the selected factors.
6. Select the experimental design.
7. Have a plan to control extraneous variables.
8. Perform the experiment according to the design.
9. Analyze and draw conclusions.
10. Verify and document the new process.
11. Propose the next study.

An experiment with two factors is presented to start the process of defining terms and procedures. The techniques developed in this small experiment can then be used in experiments with many more factors.

Example 1: Bond Strength

A manufacturer of laminated papers takes two rolls of kraft paper and extrudes a layer of polymer in between, simultaneously pressing the three components into a *sandwich.* The customer has complained about the lack of adherence of the polymer to the two paper layers. Therefore, the objective of the study is to maximize the bond strength. Note that the objective must be defined up front. The manufacturer has developed a test for this bonding using a Bond Meter. (Don't look for this in your equipment catalog!) This is the key output measure; in DOE, it is called the *response variable.* The team working on improving the bond has decided that the best candidate *factors* for controlling bond strength are polymer temperature and paper source. *Factors* are process variables that can be controlled at will during the experiment. Think of them as *knobs* that you can turn as you wish. (Such factors are also referred to as *independent variables* in some references.) These factors were selected based on the review of data from the process and on the brainstorming by the team. After much discussion, the *levels*—the different options or settings for each factor in the experiment—were set as shown below.

Factor	Low Level	High Level
A. Polymer Temperature, °F	580	600
B. Paper Source	Vendor *Y*	Vendor *X*

Note that factor *A* is a *quantitative* factor since its levels can be set along a relatively continuous measurement scale. Factor *B* is *qualitative* since its levels are discrete; i.e., there are a finite number of levels available. Where possible, quantitative factors are preferred since that permits the experimenter to follow such a factor to an optimum condition with respect to the response.

Next, the *experimental design* must be defined. The experimental design is the definition of the collection of trials to be run in the experiment. In this example, the design is all possible combinations of the chosen factor levels, called a *full-factorial* design. Since each factor has two levels, this is a 2^2 design requiring four unique *treatments* or *runs*. (The base 2 refers to the number of levels; the exponent refers to the number of factors.) These four runs are the four combinations of the levels of the factors using either the labels *high* and *low* or plus and minus signs to identify the actual levels:

1. A low; B low = A_-B_-
2. A low; B high = A_-B_+
3. A high; B low = A_+B_-
4. A high; B high = A_+B_+

A *run* or *treatment* is the unique combination of the factor levels. Note that each run may be performed more than once. The procedure of performing more than one trial of each run is referred to as *replication* when each experimental trial utilizes a completely new setup. A *replicate* is an independent and random application of the run, including the setup. This is considerably different from a *repeat*, which is a repetition of a run without going through a new setup. The 2^2 design defines four treatments or runs. To reduce the impact of the inherent variation in the process, each run is replicated for a total of eight trials. These eight trials must be carried out in a random order to minimize the risk of bias in the results due to unknown or uncontrolled factors.

Randomization refers to the order in which the trials of an experiment are performed. Randomization can be achieved by numbering the trials and then drawing numbers from a hat, using a table of random numbers, shuffling numbered cards, etc. This is important to protect against uncontrolled and/or unknown influences of variables that are not part of the experiment. As an example, assume that an experiment has a single factor, pressure. Assume also that, unknown to the experimenters, a thermostat reading is drifting steadily downward and that the background temperature affects the response. If all the low-pressure trials are performed on day one and the high-pressure trials on day two, is their difference due to the change in pressure or to the change in temperature? The experimenters cannot be sure and, in fact, may not even know there is a problem. If the temperature impact is major, one could erroneously conclude that pressure is a causative factor. In order to minimize this risk of unknown influence, the experimenters randomly assign the order of testing to improve the chances of averaging out this *bias* or *distortion* of the responses related to the factor(s) under study.

There are times one cannot randomize due to physical or cost constraints. Such cases generally lead to *blocking* of the experiment into sections defined by the factor that cannot be randomized. For instance, there may not be enough material from one batch of raw material to complete the experiment. Instead, one may *block* by carrying out a carefully defined set of trials with one batch and the remaining trials with a second batch. Such blocking minimizes the risk of the nuisance-factor batches creating excessive estimates of the inherent variation.

This first part of Chapter 2 is the *design* part of DOE: factor selection, setting levels, defining treatments, randomization of the order of performance. Now the eight trials must be performed exactly as required, keeping all other factors constant. The responses or results that were obtained are shown in Table 2.1.

TABLE 2.1. Data for Example 1.

		A. Polymer Temperature, °F	
		580 (Low) –	600 (High) +
B. Vendors	Vendor Y (Low) –	18.6 17.4 $\overline{Y} = 18.00$	17.5 16.5 $\overline{Y} = 17.00$
	Vendor X (High) +	18.2 16.7 $\overline{Y} = 17.45$	22.9 22.2 $\overline{Y} = 22.55$

Notice the use of plus and minus signs as another way to indicate the two levels. Analysis of these data will be done using the eight-step analytical procedure that follows. These eight steps will work for experiments with any number of factors, as long as the factors have two levels. At first glance, the limitation to two levels may seem very restrictive. In actuality, it permits a remarkable efficiency in the number of trials needed. The primary penalty is that the relationship between a response and a factor with two levels is assumed to be linear. Techniques for identifying nonlinear relationships (curvature) are presented later.

The Eight Steps for Analysis of Effects

1. Calculate effects.
2. Make a Pareto chart of effects.
3. Calculate the standard deviation of the experiment, S_e.
4. Calculate the standard deviation of the effects, S_{eff}.
5. Determine the t-statistic.
6. Calculate the decision limits and determine the significant effects.
7. Graph significant effects.
8. Model the significant effects.

The Analytical Procedure

1. Calculate Effects.

Main effects are defined as the difference in the average response between the high and low levels of a factor. The *Effect of A* is written as *E(A)*. Using plus and minus signs to represent high and low levels of a factor, main effects are defined as

$$E(A) = \overline{Y}_{A+} - \overline{Y}_{A-} = \frac{22.55 + 17.00}{2} - \frac{17.45 + 18.00}{2} = 19.78 - 17.73 = 2.05$$

The main effect of +2.05 means that the average bond strength at the high level of temperature (600°F) was 2.05 units higher than the average bond strength at the low temperature (580°F). The same data can now be used to determine the main effect of factor *B* (vendor) on the process. Note that a plus sign is used to designate vendor *X* while a minus sign is used to designate vendor *Y*. This is an arbitrary choice that once chosen must be used consistently.

$$E(B) = \overline{Y}_{B+} - \overline{Y}_{B-} = \frac{22.55 + 17.45}{2} - \frac{17.00 + 18.00}{2} = 20.00 - 17.50 = 2.50$$

This indicates that the paper supplied by vendor *X* averaged 2.50 units higher in bond strength than that supplied by vendor *Y*. *An effect is the difference in averages.* Note that in a DOE the same data are used for more than one factor. This is the fundamental concept of DOE: an experiment is used to define many effects, rather than performing an experiment for each factor. The graphs in Figures 2.1 and 2.2 indicate the direction of influence for the factors and communicate the meaning of the effects very clearly.

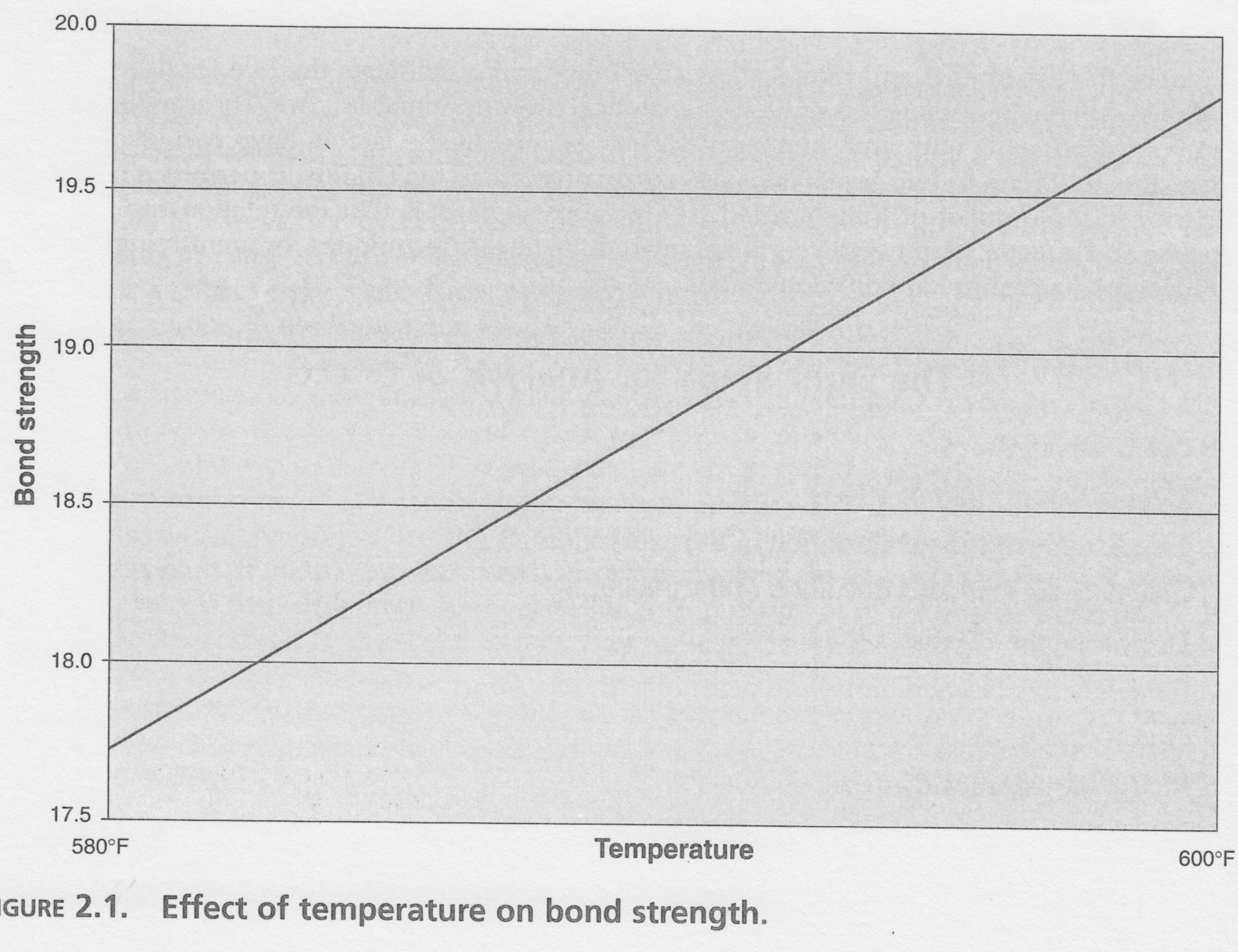


FIGURE 2.1. Effect of temperature on bond strength.

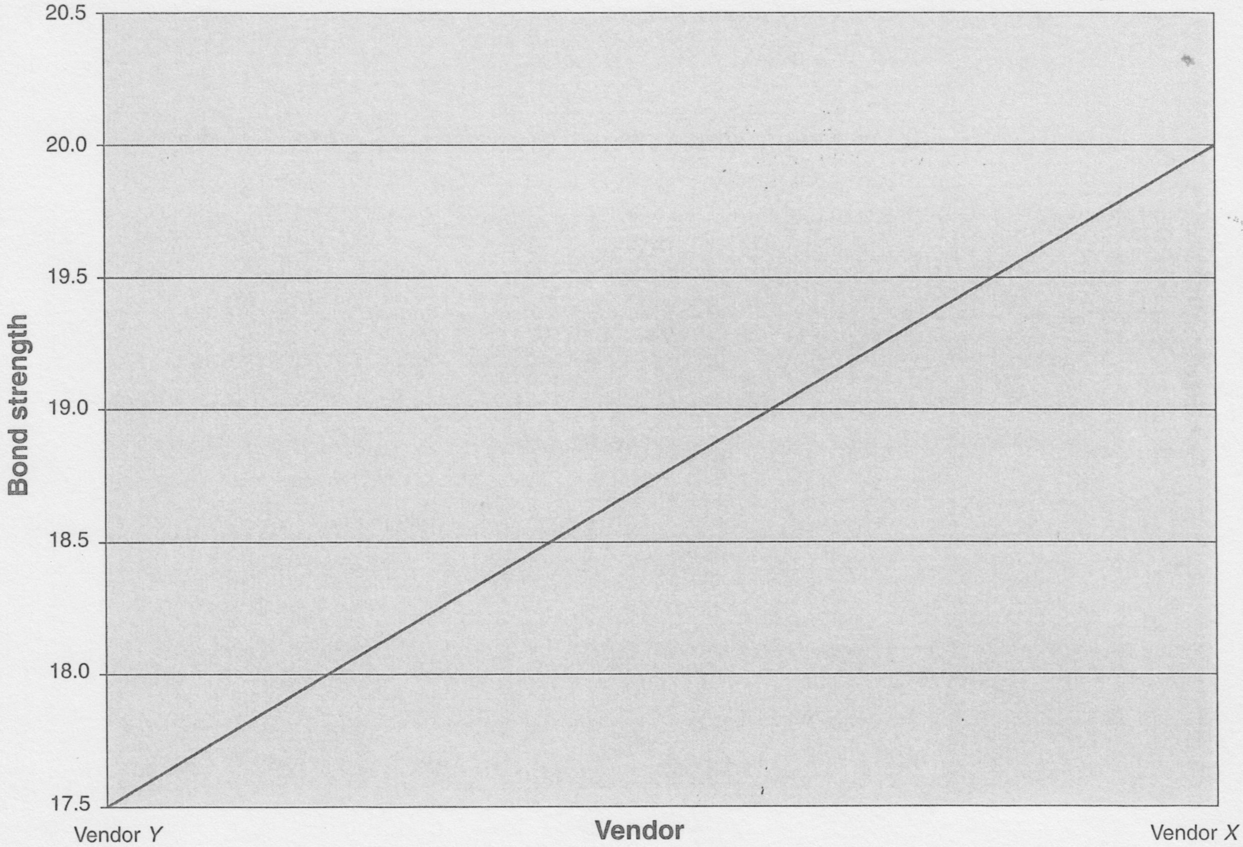


FIGURE 2.2. Effect of vendor on bond strength.

The plots of the main effects of temperature and vendor indicate that a higher temperature results in a stronger bond and that vendor X is superior to vendor Y. Observe the data in Table 2.1. Are those results *always* true and to the amount expected? Are there certain combinations of temperature and vendor that don't seem to provide the expected results? An *interaction* occurs when a particular combination of two factors does something unexpected from simply observing their main effects. An interaction is defined as one-half of the difference between the effect of A at the high level of B and the effect of A at the low level of B. Mathematically, this is

$$E(AB) = \tfrac{1}{2}[(\overline{Y}_{A+} - \overline{Y}_{A-})_{B+} - (\overline{Y}_{A+} - \overline{Y}_{A-})_{B-}]$$

Using the data in the table,

$$\text{E(AB)} = \tfrac{1}{2}[(22.55 - 17.45) - (17.0 - 18.0)] = \tfrac{1}{2}[5.1 - (-1.0)] = 3.05$$

The interaction effect essentially increases or decreases the main effect in the experiment by 3.05 units of strength. For example, the main effect of temperature equals +2.05. However, when paper from vendor X is used, the effect of temperature is actually 5.10; i.e., (22.55 – 17.45). Conversely, when paper from vendor Y is used, the temperature effect is –1.0; i.e., (17.0 – 18.0). This means that when using paper from vendor Y, higher bond strengths are achieved with *lower* temperature. The plot reveals what the interaction effect means (see Figures 2.3 and 2.4 on page 10).

Interactions can be plotted as *vendor-temperature* or as *temperature-vendor;* both are correct and are simply two views of the same phenomenon. The two plots do *not* always look alike. If in doubt about which is more useful, plot both!

A final precaution on interactions: if an interaction proves to be significant, the interaction chart is more important than the main effect charts. If there is an interaction, the main effect describes average results whereas the interaction is more appropriate in describing the joint effect.

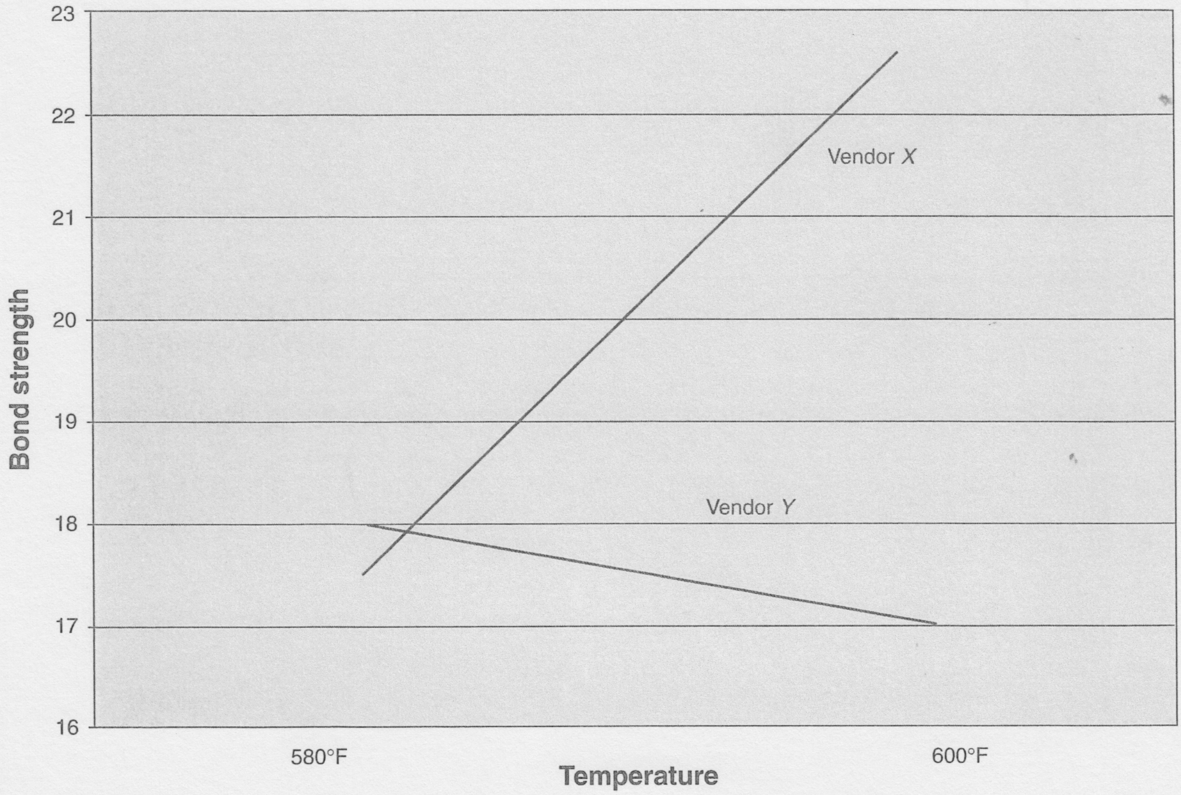


FIGURE 2.3. Temperature-vendor interaction.

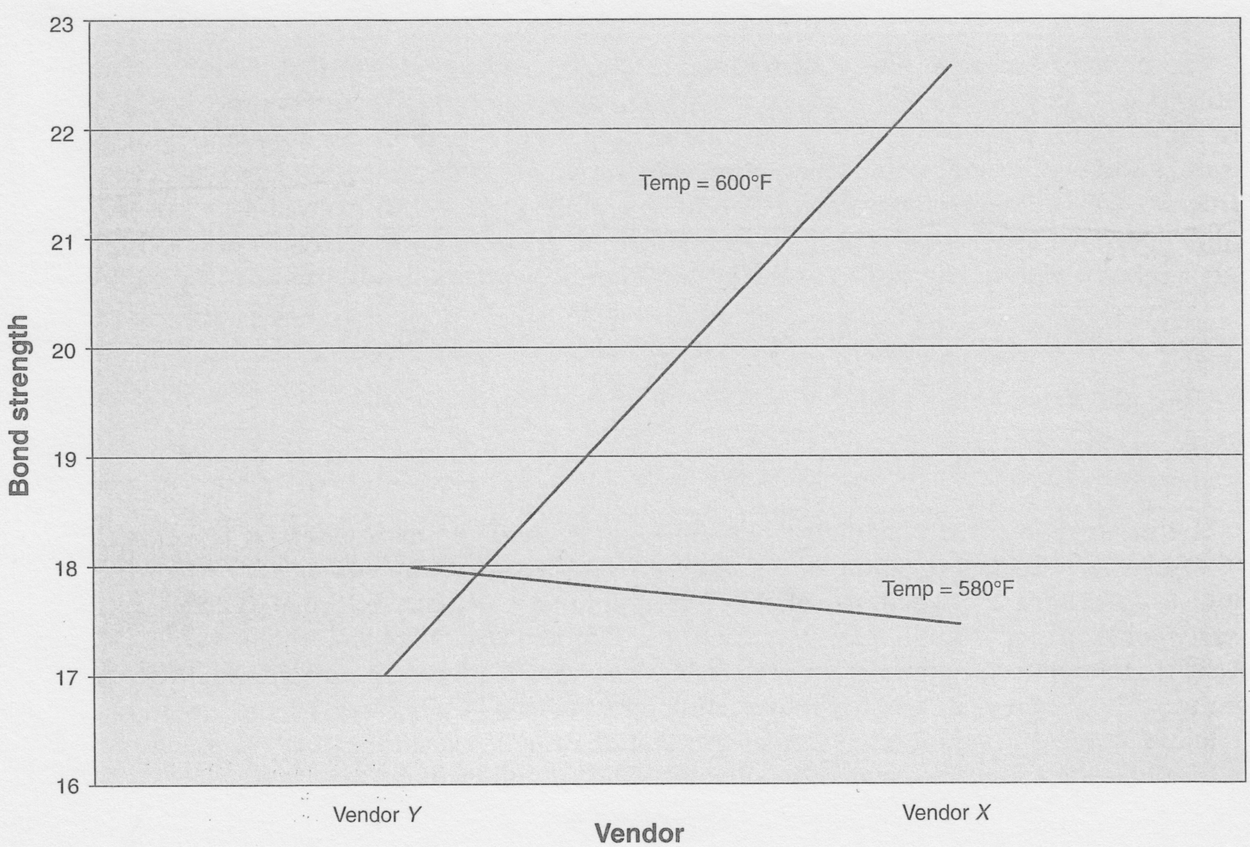


FIGURE 2.4. Vendor-temperature interaction.

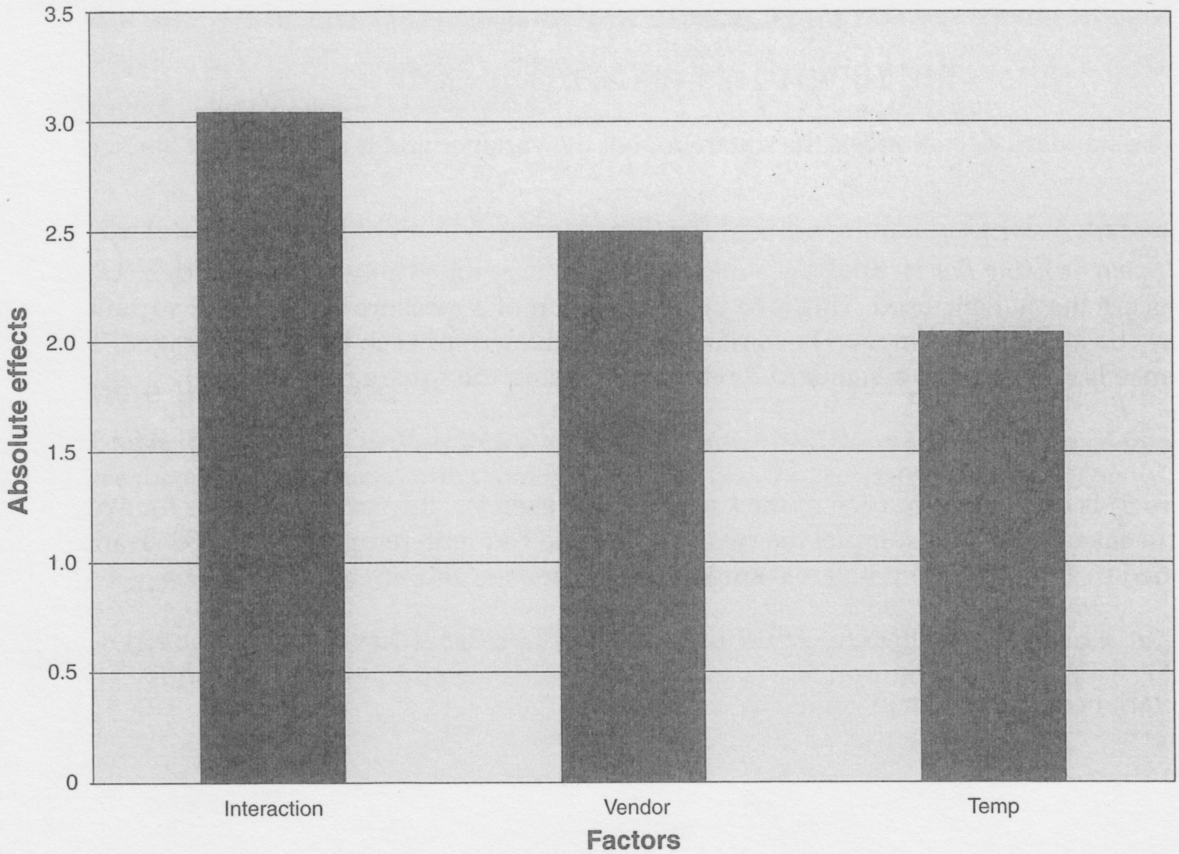


FIGURE 2.5. Pareto chart of effects.

2. Make a Pareto Chart of Effects.

To show the relative importance of the effects, plot their absolute value as a Pareto chart (see Figure 2.5). The Pareto chart, which becomes more useful when larger experiments are studied, is often sufficient to determine what effects are meaningful.

The problem now is to test these effects for significance; i.e., how do we show that the results of the experiment are beyond the inherent variation to be expected in the experiment? The next four steps will determine whether these three effects are significant or meaningful.

3. Calculate the Standard Deviation of the Experiment, S_e.

Determination of significance requires calculation of a *standard deviation* as a measure of the *inherent variation* or *experimental error* in the process. In SPC, this was referred to as *sigma* and used the symbol σ_e. In DOE, the symbol S_e is used. To obtain S_e, the *variance* (S^2) is calculated for each run or treatment. (Here is where a scientific calculator will be worth its cost! Just be sure to use the key for a divisor of n–1.) These variances are then averaged and converted to a standard deviation (S_e) by taking the square root.

Let's have a tutorial on this. A *variance* is the square of the deviation of each observation of a sample from the sample average. The formula is

$$S^2 = \frac{\Sigma(X_i - \overline{X})^2}{n-1}$$

Assume three responses: 12, 14, 16. Since the average is 14,

$$S^2 = \{(12-14)^2 + (14-14)^2 + (16-14)^2\}/(3-1) = \{(-2)^2 + 0^2 + 2^2)\}/2 = 4.0$$

The standard deviation S is the square root of the variance and is a measure of inherent variation.

$$S = \sqrt{S^2} = \sqrt{4} = 2.0.$$

Don't plan on doing this by hand; use a calculator! Variances are averaged even though standard deviations are the statistic used. This is to permit creation of a measure of spread or variation that will always be a positive number. The variances are calculated for each run and averaged. The average variance is converted to a standard deviation by taking the square root:

$$S_e = \sqrt{(\Sigma S_i^2 / k)}$$

where S_i^2 is calculated for each of the k runs. In this exercise, the variances of the four runs of Table 2.1 are calculated. For example, the two trials for the run with temperature = 580° F and vendor Y resulted in 18.6 and 17.4 with an average of 18.0. Then

$$S_i^2 = [(18.6 - 18.0)^2 + (17.4 - 18.0)^2]/(2-1) = (.36 + .36)/1 = .72$$

The variances are shown in Table 2.2.

TABLE 2.2. Variances for each run of example 1.

	Temperature	
	580°F	600°F
Vendor Y	.720	.500
Vendor X	1.125	.245

$$S_e^2 = (.72 + .50 + 1.125 + .245)/4 = 0.648$$

and

$$S_e = \sqrt{.648} = .80$$

This estimate of the inherent variation represents (1) measurement variation, plus (2) the inability of the experimenter to repeat the conditions, plus (3) the inability of the process to repeat the same response for the same conditions.

There is more than one way to calculate the standard deviation of the experiment. This technique is used since it is the classic approach and more closely matches what will later be seen in software. Range techniques from SPC could also be used with little loss in precision of the estimate, that is,

$$S_e = \frac{\overline{R}}{d_2}$$

where $\overline{R}$ is the average of the ranges of the runs and d_2 is a tabular value based on the number of replicates.

4. Calculate the Standard Deviation of the Effects, S_{eff}.

Effects are differences between averages, requiring a modified standard deviation. This is called the *standard deviation of the effects* and is defined as

$$S_{eff} = S_e \sqrt{(4/N)}$$

where N is the total number of trials. This formula will hold for any number of trials as long as the factors have two levels. In the example,

$$S_{eff} = .80\sqrt{(4/8)} = 0.57$$

5. Determine the t-Statistic.

To use the t-table, the *degrees of freedom* in the experiment must be determined. Degrees of freedom (*d.f.*) measure the amount of information available to estimate the standard deviation. The calculation is

$$d.f. = (\text{\# of observations per run} - 1) \times (\text{\# of runs}) = (2 - 1) \times (4) = 4$$

Next, a reference number must be selected from the t-table on page 66 in the appendix. Ninety-five percent confidence is customarily considered necessary to claim significance in effects. Confidence is

$$1 - \textit{alpha risk}$$

where the alpha risk is the chance of erroneously claiming significance. The typical alpha risk (also written with the Greek symbol α) is therefore 5 percent. This table is based on that level of confidence. Referring to the t-table for 4 degrees of freedom and $\alpha = 5$ percent, the tabular value is $t = 2.78$.

6. Calculate the Decision Limits and Determine the Effects.

The *decision limits (DL)* for the significance of effects in this DOE can now be calculated. The question is whether the computed effects are significantly different from zero and, therefore, not due to random variation. If the effects are outside the zone defined by the decision limits, the effects are considered real or significant. The decision limits are calculated by

$$DL = \pm\,(t_{\alpha,df})\,(\sigma_{eff}) = \pm\,(2.78)\,(.57) = \pm\,1.58$$

The results are shown graphically in Figure 2.6.

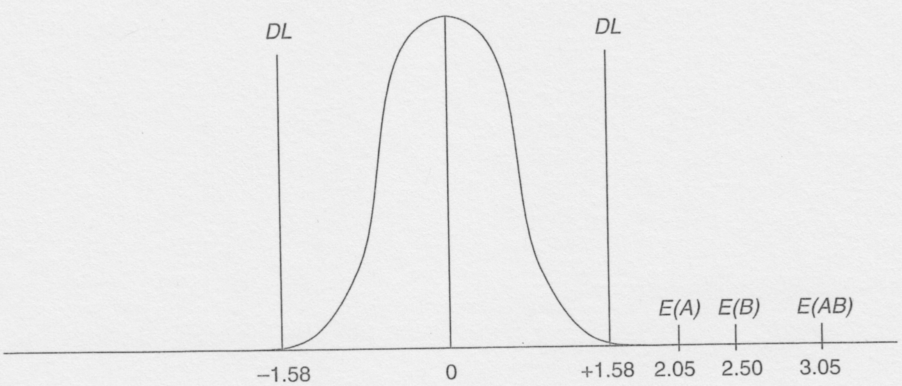


FIGURE 2.6. Decision limits.

What is the conclusion here? All three of the effects in the example exceed the limits and are judged to be real or significant. Since the interaction effect was significant, one must jointly consider the two factors to determine how to optimize the process. Since the objective was to maximize bond strength, one should use paper from vendor X and operate at a temperature of 600°F. Refer to Figures 2.3 and 2.4.

7. Graph Significant Effects.

Actually, graphs need only be made for significant effects. If an effect is not significant, the graph is of no interest. If an interaction is significant, that graph is more meaningful than those of the individual factors since an interaction means that the experimenter must consider the factors jointly. The graphs for this example have already been shown in Figures 2.1, 2.2, 2.3, and 2.4.

8. Model the Significant Effects.

The *model* or *prediction equation* is useful to predict the optimum outcome for future validation experiments. The model is a linear equation of the following form, using only the significant effects. However, if an interaction effect is significant, the terms for the two main effects are also included *even if they are not significant.* This is due to the *hierarchy rule* for defining a model. This rule becomes more important in advanced techniques for optimization. The levels of *A*, *B*, and *AB* are coded by –1 for the low levels and +1 for the high levels. The term $\bar{\bar{Y}}$ represents the average of all the data.

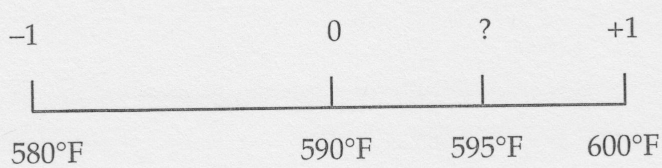
$$\hat{Y} = \bar{\bar{Y}} + \frac{E(A)}{2}A + \frac{E(B)}{2}B + \frac{E(AB)}{2}AB$$

Since our objective is to maximize bond strength, *A* and *B* are set at the high (+) levels that will provide maximum strength. Specifically,

$$\hat{Y} = 18.75 + (2.05/2)\,(+1) + (2.50/2)(+1) + (3.05/2)\,(+1)(+1) = 22.55$$

If the best combination of temperature and vendor paper is used, bond strength is estimated to average 22.55. This is better than the results predicted simply from the main effects of *A* and *B* due to the contribution of the interaction term. An interaction either increases the response over that expected due to main effects alone or it reduces the response below that expected from main effects.

What if one needed to know the expected response for a temperature of 595°F with paper from vendor X? The model uses coded values for the levels so the temperature of 595°F must be changed to coded values in order to interpolate the results:



Since 1 unit = 10° and 595° is 5 degrees from the zero point equivalent to 590°, the coded equivalent of 595° is 5/10 = + .5. This value is then inserted into the model for *A* with *B* = +1:

$$\begin{aligned}\hat{Y} &= \bar{\bar{Y}} + \frac{E(A)}{2}A + \frac{E(B)}{2}B + \frac{E(AB)}{2}AB \\ &= 18.75 + (2.05/2)A + (2.50/2)B + (3.05/2)\,AB \\ &= 18.75 + 1.03A + 1.25B + 1.53AB \\ &= 18.75 + 1.03(.5) + 1.25(1) + 1.53(.5)(1) = 21.28\end{aligned}$$

This interpolation procedure is useful when there are only a couple of significant factors. Otherwise, too many assumptions must be made since one would have a single equation in several unknowns.

What further study can be recommended if even higher bond strength is desired? Does the temperature effect continue beyond 600°F? It is dangerous to extrapolate beyond the range of study (580° to 600°F). Plan a verification study at 600°F and at 610°F with paper from vendor *X*. This will verify the previous conclusions and explore the potential for even better performance.

The objective in this example was to achieve stronger bond strength. Suppose the objective had been different. What if the bond strength problem was not one of inadequate strength but of too much variation in the strength since temperature could not be tightly controlled? What conclusions would be drawn in that circumstance? What conditions would make the process *robust* or insensitive to changes in the process variables of temperature and vendor? The interaction plots show that changes in temperature have little effect on bond strength if paper from vendor *Y* is used. This is why the objective of the study must be clearly understood before the experiment is finalized.

For demonstration purposes, again assume a different objective. Suppose that, instead of achieving a maximum or minimum, the experimenter needed to hit a target response of 20. This would lead one to determine the setting of the quantitative (numeric) factor of temperature. Based on the interaction graph (Figure 2.4), note that only vendor *X* material provides a range of responses that includes the value of 20. If *B* is set at +1 for vendor *X*, what setting for factor *A* (temperature) would provide a forecast of the target value? To determine this, solve the model for *A* where $B = +1$ (vendor X) and $Y = 20$ (target response):

$$\hat{Y} = \bar{\bar{Y}} + \frac{E(A)}{2}A + \frac{E(B)}{2}B + \frac{E(AB)}{2}AB$$

$$20 = 18.75 + 1.03A + 1.25B + 1.53\,AB$$

$$20 = 18.75 + 1.03A + 1.25 + 1.53\,A$$

$$2.56A = 0$$

$A = 0$ in coded units or, by interpolation, A = 590° in actual units.

This completes the analytical procedure. It is very important that the conclusions from this experiment be verified either by operating at the recommended conditions or by further experimentation. (More advanced techniques are covered in the references on page 63 to more completely address the subject of optimization.) All reports, graphs, and recommendations should be put in layman's terms, not in the statistical terminology that is used in the analysis. Remember that the average employer may not know what *A*– means, but he or she understands Pareto charts (the 80/20 rule) and graphs as long as the references are in familiar language. A box with the steps of the analytical procedure is provided for quick reference.

THE EIGHT STEPS FOR ANALYSIS OF EFFECTS

1. Calculate effects.
2. Make a Pareto chart of effects.
3. Calculate the standard deviation of the experiment, S_e.
4. Calculate the standard deviation of the effects, S_{eff}.
5. Determine the t-statistic.
6. Calculate the decision limits and determine the effects.
7. Graph significant effects.
8. Model the significant effects.

These steps are the fundamental analysis techniques for any number of factors at two levels. Remember that small DOEs such as the previous example are usually used in refining experiments after larger screening experiments or when some quick troubleshooting is needed. The spacing of the quantitative levels (temperature in our example) is also important. In a screening design, the levels should be set as wide or as bold as practical to make it easier to discover significant factors for later study. In a refining design, the levels would be set more closely and more replication would be required since the size of the effects would be reduced. Additional data would be needed to demonstrate significance of the smaller effects.

A final caution is needed on the statistical control of the process. Lack of control increases the experimental error and can also create false effects. It is important to stabilize the process, i.e., eliminate the special causes. Otherwise, replication must be increased to overcome the distortion (overestimate) of the estimate of inherent variation (S_e) due to the presence of special causes of variation.

Review of the Experimental Procedure

Having completed an experiment and gone through the analytical procedures, one can review the broader concepts of experimentation. There are a series of *logic steps* that must be addressed as one prepares to launch a DOE.

1. What is the process to be studied? How broadly or narrowly is it defined? A flowchart is a good tool for this analysis.
2. What is the response? What needs to be improved? Should there be more than one response? Note that additional responses are free, costing only the measurements!
3. What is the measurement precision? Is there bias in the measurement system? Has a measurement analysis been completed? Is it adequate?
4. Generate candidate factors. This is best done with a small team using brainstorming after a review of all available data and information on the process and response variables. A cause-and-effect diagram with a flowchart of the process is useful with this brainstorming. The trick is to be innovative, to think outside usual boundaries, and yet not try to reinvent proven technology. Provide opportunities for surprises! The team should be knowledgeable about the issues and follow the rules for brainstorming.
5. Determine the levels for the factors selected for the DOE. In screening experiments, the rule is to have levels broadly spaced but not to the point of being foolhardy. In refining experiments, levels will be much tighter and will require more replication.
6. Select the experimental design. This is the set of treatments or runs that will be performed. This also includes deciding on the amount of replication. Finally, the randomized order of the trials is determined. (*Randomization* is the insurance policy against misleading conclusions due to outside influence during the experiment.)
7. Establish a plan to control (or at least monitor) extraneous variables.
8. Perform the experiment according to the design. The DOE *must* be carried out per its design. Identify trial materials carefully. Keep good notes.
9. Analyze, draw conclusions, and assess process impact. What process variables can be changed—and how—to improve the process?
10. Verify and document the new process as defined by the experiment.
11. Propose the next study for continuation of this project, or declare the project complete. Make sure that all reports that go beyond the team are in language and terminology that are easily understood.