

$$\text{Tool life} = 106.90 - 1.3614(\text{cutting speed})$$

This equation can be used to predict tool life by plugging in values of cutting speed. Extreme caution should be used in making predictions outside the actual sample space (e.g., for cutting speeds above or below the tested maximum or minimum), however, as these are tenuous without confirmation by observation.

Although a prediction equation can be found mathematically, it should not be used without knowing how "good" it is. A number of criteria exist for judging the adequacy of the prediction equation. One common measure is  $R^2$ , the proportion of variation explained by the prediction equation.  $R^2$ , or the coefficient of determination, is the ratio of the variation due to the regression to the total variation. The higher  $R^2$ , the greater the probable utility of the prediction equation in estimating  $Y$  based on  $X$ .

Another measure of the degree of association between two variables is the simple linear correlation coefficient,  $r$ . This is the square root of the coefficient of determination, so that the values of  $r$  range from  $-1$  to  $+1$ . A positive  $r$  is consistent with a positive relationship (an increase in one variable is associated with an increase in the other), whereas the opposite is true of a negative  $r$  (an increase in one variable is associated with a decrease in the other). Scatter plots are strongly recommended when interpreting correlations, especially as very different patterns can result in identical values of  $r$ . The significance level of  $r$  varies with sample size; statistical software is recommended to obtain exact significance levels.

The above discussion introduces simple linear correlation and regression—the direction and strength of a relationship between two variables, or prediction of a dependent variable,  $Y$ , from a single predictor variable,  $X$ . A natural extension of this is multiple regression that allows for two or more independent variables. For a discussion of how to estimate and examine a multiple regression prediction equation, see Kutner et al. (2004).

## 20.7.6. Analysis of Variance

Analysis of Variance (ANOVA) is an approach related to linear regression, falling into the class of what are called general linear models. However, unlike regression, the  $X$  is discrete rather than continuous (noting that general linear models actually can blend characteristics of both regression and ANOVA). In ANOVA, the total variation of all measurements around the overall mean is divided into sources of variation that are then analyzed for statistical significance. It is used in situations where the investigator is interested in comparing the means among two or more discrete groups. For example, an investigator may be interested in comparing performance among three different machine configurations. The ANOVA analysis detects a difference somewhere among the means (i.e., at least one mean is different from the others), and confidence intervals or follow-up tests such as pairwise comparisons can be applied to determine which mean (or means) is different. ANOVA is the basis for design of experiments, discussed next.

## 20.7.7. Design of Experiments

With origins in the pioneering work in agriculture of Sir Ronald A. Fisher, designed experiments have taken on an increasingly significant role in quality improvement in the business world. This section will first compare the classical and designed approaches to experimentation, thereby providing the reader with an understanding as to the limitations of traditional methods and the power of contemporary methods. Next, basic concepts and terminology will be introduced in the context of an example improvement problem, followed by an overview of different types of designs and the typical progression through a series of designed experiments. The section finishes with the related topic of Taguchi designs.

## 20.7.7.1. Contrast between the Classical and Contemporary Methods of Experimentation

The classical method of experimentation is to vary one factor at a time (sometimes called OFAT), holding everything else constant. By way of example, and to illustrate the need for designed experiments, consider the case of a certain fellow who decided he wanted to investigate the causes of intoxication. As the story goes, he drank some whiskey and water on Monday and became highly inebriated. The next day, he repeated the experiment holding all variables constant except one... he decided to replace the whiskey with vodka. As you may guess, the result was drunkenness. On the third day, he repeated the experiment for the last time. On this trial, he used bourbon in lieu of the whiskey and vodka. This time it took him two days just to be able to gather enough of his faculties to analyze the experimental results. After recovering, he concluded that water causes intoxication. Why? Because it was the common variable!

The contrast between this traditional method and the designed approach is striking. In particular, a designed approach permits the greatest information to be gained from the fewest data points (efficient experimentation), and allows the estimation of interaction effects among factors. [Table 20.11](#) compares these two approaches in more detail for an experiment in which there are two factors (or variables) whose effects on a characteristic are being investigated (the same conclusions hold for an experiment with more than two factors).

**Table 20.11** Comparison of Classical and Modern Methods of Experimentation

Criteria		Classical	Modern
<i>Source: Quality Planning and Analysis, Copyright 2007. Used by permission.</i>			
Basic procedure		Hold everything constant except the factor under investigation. Vary that factor and note the effect on the characteristic of concern. To investigate a second factor, conduct a separate experiment in the same manner.	Plan the experiment to evaluate both factors in one main experiment. Include in the design measurements to evaluate the effect of varying both factors simultaneously.
Experimental conditions		Care should be taken to have material, workers, and machine constant throughout the entire experiment.	Realizes difficulty of holding conditions reasonably constant throughout an entire experiment. Instead, experiment is divided into several groups or blocks of measurements. Within each block, conditions must be reasonably constant (except for deliberate variation to investigate a factor).
Experimental error		Recognized but not stated in quantitative terms.	Stated in quantitative terms.
Basis of evaluation		Effect due to a factor is evaluated with only a vague knowledge of the amount of experimental error.	Effect due to a factor is evaluated by comparing variation due to that factor with the quantitative measure of an experimental error.
Possible bias due to sequence of measurements		Often assumed that sequence has no effect.	Guarded against by randomization.
Effect of varying both factors simultaneously ("interaction")		Not adequately planned into experiment. Frequently assumed that the effect of varying factor 1 (when factor 2 is held constant at some value) would be the same for any value of factor 2.	Experiment can be planned to include an investigation for interaction between factors.
Validity of results		Misleading and erroneous if interaction exists and is not realized.	Even if interaction exists, a valid evaluation of the main factors can be made.
Number of measurements		For a given amount of useful and valid information, more measurements are needed than in the modern approach.	Fewer measurements needed for useful and valid information.
Definition of problem		Objective of experiment frequently not defined as necessary.	Designing the experiment requires defining the objective in detail (how large an effect do we want to determine, what numerical risks can be taken, etc.).
Application of conclusions		Sometimes disputed as applicable only to the controlled conditions under which the experiment was conducted.	Broad conditions can be planned in the experiment, thereby making conclusions applicable to a wider range of actual conditions.

## 20.7.7.2. Concepts and Terminology—An Example Designed Experiment

Suppose that three detergents (A, B, C) are to be compared for their ability to clean clothes in an automatic washing machine. The "whiteness" readings obtained by a special measuring procedure are the dependent, or response, variable. The independent variable under investigation (detergent) is a factor, and each variation of the factor is called a level; in this case, there are three levels. A treatment is a single level assigned to a single factor, detergent A. A treatment combination is the set of levels for all factors in a given experimental run. A factor may be qualitative (different detergents) or quantitative (water temperature). Finally, some experiments have a fixed-effects model (i.e., the levels investigated represent all levels of concern to the investigator—for example, three specific washing machines or brands). Other experiments have a random effects model, that is, the levels chosen are just a sample from a larger population (e.g., three operators of washing machines). A mixed-effects model has both fixed and random factors.

**Figure 20.16** outlines six possible designs of experiments, starting with the classical design in (a). Here, all factors except detergent are held constant. Thus, nine tests are run, three with each detergent with the washing time, make of machine, water temperature, and all other factors held constant. One drawback of this design is that the conclusions about detergent brands apply only to the specific conditions of the experiment.

**Figure 20.16** Some experimental designs. (*Quality Planning and Analysis*, Copyright 2007. Used by permission.)

A	B	C	
-	-	-	
-	-	-	
-	-	-	
(a)			
I	II	III	
A	B	C	
A	B	C	
A	B	C	
(b)			
I	II	III	
C	B	B	
A	C	B	
A	A	C	
(c)			
I	II	III	
B	A	C	
C	C	A	
A	B	B	
(d)			
	I	II	III
1	C	A	B
2	B	C	A
3	A	B	C
(e)			
I	II	III	
ABC	ABC	ABC	
1 ---	---	---	
2 ---	---	---	
3 ---	---	---	
(f)			

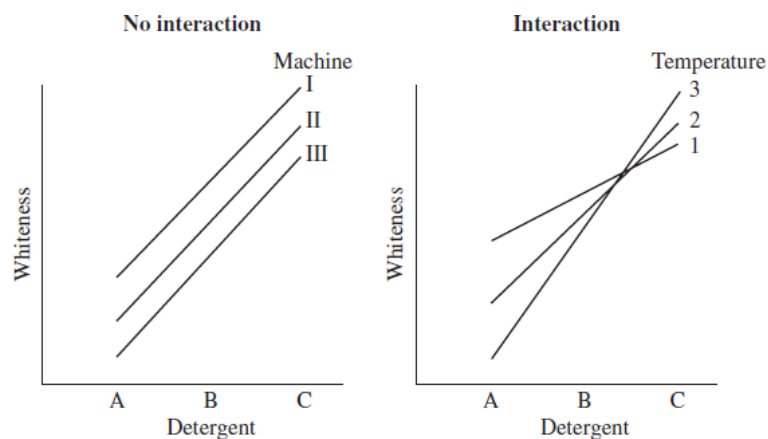
Design (b) recognizes a second factor at three levels (i.e., washing machines brands I, II, and III). However, in this design, it would not be known whether an observed difference was due to detergents or washing machine (they are said to be confounded).

In design (c), the nine tests are assigned completely at random, thus the name "completely randomized design." However, detergent A is not used with machine brand III, and detergent B is not used with machine brand I, thus complicating the conclusions.

Design (d) shows a randomized block design. Here each block is a machine brand, and the detergents are run in random order within each block. This design guards against any possible bias due to the order in which the detergents are used and has advantages in the subsequent data analysis and conclusions. First, a test of hypothesis can be run to compare detergents and a separate test of hypothesis run to compare machines; all nine observations are used in both tests. Second, the conclusions concerning detergents apply for the three machines and vice versa, thus providing conclusions over a wider range of conditions.

Now suppose that another factor such as water temperature is also to be studied, using the Latin square design shown in (e). Note that this design requires using each detergent only once with each machine and only once with each temperature. Thus, three factors can be evaluated (by three separate tests of hypothesis) with only nine observations. However, there is a danger. This design assumes no interaction among the factors. No interaction between detergent and machine means that the effect of changing from detergent A to B to C does not depend on which machine is used, and similarly for the other combinations of factors. The concept of interaction is shown in Fig. 20.17. There is no interaction among the detergents and the machines. But the detergents do interact with temperature. At high temperatures, C is the best performer. At low temperatures, A performs best.

**Figure 20.17** Interaction. (Quality Planning and Analysis, Copyright 2007. Used by permission)



Finally, the main factors and possible interactions could be investigated by the factorial design in design (f) in Fig. 20.16. Factorial means that at least one test is run for every combination of main factors, in this case  $3 \times 3 \times 3$  or 27 combinations. Separate tests of hypothesis can be run to evaluate the main factors and also possible interactions. Again, all the observations contribute to each comparison. When there are many factors, a portion of the complete factorial (i.e., a "fractional factorial") is useful when experimental resources are limited (see its application in a sequential testing approach, below).

Most problems can be handled with one of the standard experimental designs or a series of these. Designs can be classified by the number of factors to be investigated, the structure of the experimental design, and the kind of information the experiment is intended to provide (Table 20.12). For a description of both the design and analysis of various design structures, see Box et al. (2005). Another excellent general reference is Myers et al. (2009) for a detailed look at response surface designs.

**Table 20.12** Classification of Designs

Design	Type of Application
Source: Adapted from JQH5, Table 47.3.	
Completely randomized	Appropriate when only one experimental factor is being investigated
Factorial	Appropriate when several factors are being investigated at two or more levels and interaction of factors may be significant
Blocked factorial	Appropriate when number of runs required for factorial is too large to be carried out under homogeneous conditions
Fractional factorial	Appropriate when many factors and levels exist and running all combinations is impractical
Randomized block	Appropriate when one factor is being investigated and experimental material or environment can be divided into blocks or homogeneous groups
Balanced incomplete block	Appropriate when all the treatments cannot be accommodated in a block
Partially balanced incomplete block	Appropriate if a balanced incomplete block requires a larger number of blocks than is practical
Latin square	Appropriate when one primary factor is under investigation and results may be affected by two other experimental variables or by two sources of nonhomogeneity. It is assumed that no interactions exist.
Youden square	Same as Latin square, but number of rows, columns, and treatments need not be the same
Nested	Appropriate when objective is to study relative variability instead of mean effect of sources of variation (e.g., variance of tests on the same sample and variance of different samples)
Response surface	Objective is to provide empirical maps (contour diagrams) illustrating how factors under the experimenter's control influence the response
Mixture designs	Use when constraints are inherent (e.g., the sum of components in a paint must add to 100%)

A sequential approach to experimentation often can be helpful. Briefly, a typical sequence of designed experiments will allow an experimenter to quickly and efficiently narrow down a large number of possible factors (or X's in the  $Y = f(X)$  terminology of Lean Six Sigma) to find out which are most important, and then refine the relationships to find optimal settings for each of the vital few factors. The steps might be as follows:

1. *Screening experiment.* In this stage, a fractional factorial design may be applied that does not allow interactions to be detected, but can ferret out which of many factors have the greatest main effect.
2. *Fractional factorial design.* The smaller number of factors identified in the screening experiment are tested to allow detection of interaction effects.
3. *Full factorial design.* A small number of factors (usually no more than five) are tested to allow all main effects and higher-order (e.g., three-way, four-way) interactions to be detected and accounted for. Such designs also can detect curvature that indicates a potential optimum.
4. *Response surface design.* By adding data points in particular ways (e.g., a composite design), an experimenter can build on earlier experiments to fully characterize nonlinear relationships and pinpoint optimal settings.