jmp STATISTICAL DISCOVERY

WHITE PAPER

The Use of Designed Experimentation to Improve and Accelerate the Innovation Process

Christine M. Anderson-Cook, Los Alamos
Ryan Lekivetz, JMP Statistical Discovery LLC
Douglas C. Montgomery, Arizona State University
William H. Woodall, Virginia Tech

Contents

Innovation is required in our rapidly changing world. In this white paper, we review some key characteristics of innovation and describe the ways in which the innovation process can be aided, accelerated, and improved using designed experiments. We give some advice on how to conduct effective experiments, which can play a pivotal role in achieving both incremental and breakthrough innovations.

Introduction

The pioneering statistician George Box (shown in Figure 1) famously stated, “All experiments are designed experiments; some are poorly designed, some are well-designed.” This statement highlights the critical role of proper experimental design. Designed experimentation is widely used in many fields of science and in many industries. Some recent application areas include the study of lithium-ion batteries (Román-Ramírez et al., 2022), nanotechnology (Yuangyai et al., 2015), and the study of formulations for mRNA delivery (Hashiba et al., 2020). As seasoned experimenters, we have been associated with numerous successes in bringing design of experiments methods to diverse new areas and have seen them have dramatic impact on accelerating the innovation process. The following are a few of our favorite examples:

- Optimizing antibody microarray printing for maximum efficiency (Summers et al., 2022)
- The use of mixture-process experiments to simultaneously evaluate formulation and production process for pharmaceuticals (Anderson-Cook et al., 2004)
- Testing, evaluating, and error-checking changes in software (Lekivetz and Morgan, 2021)
- Developing crop rotations to be robust across soil types, yearly variation in growing conditions and crop pricing (Anderson-Cook et al, 2002)
- Tuning the hyperparameters of machine learning algorithms to optimize performance and robustness (Lujan-Moreno et al., 2018)
- Developmental and operational testing of complex aerospace and defense systems (Freeman et al, 2014)

- Generating a gold standard data set to evaluate the performance of radiation detection algorithms for national security (Ghawaly et al., 2020)
- Multiscale evaluation of CO2 capture process for sequesteringof greenhouse gases (Morgan et al., 2021)

We believe that design of experiments has an important role to play in driving innovation and advancement in many yet-unrealized applications.

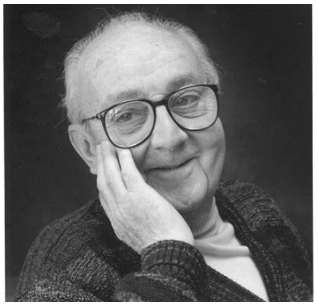


Figure 1: George E.P. Box

Experimentation can take many forms. One might think of highly controlled experiments, for example, in experimental physics, where the goal of the experiment is to support or contradict some aspect of theory. In other science and engineering applications, the goal is to understand and model fundamental patterns between inputs and outputs, allowing the exploration and exploitation of inherent relationships. In applications, the goal is more often to determine which factors affect a response of interest, culminating in the determination of the optimal operating levels of the controllable factors to create or improve a product or process.

In determining the optimal operating conditions, one could use intuition and existing knowledge to take a repeated best-guess approach until satisfactory results are obtained. Under this haphazard approach, however, learning about the process is minimal and one may never locate the optimal levels of the factors.

Another approach is the one-factor-at-a-time (OFAT) approach where the levels of each factor are varied individually while holding all other factor levels constant. This approach is terribly inefficient and typically fails to identify the optimal conditions. It is unable to provide any information about important interactions, where the effect on the response variable of changing the level of one factor depends on the level of one or more other factors. Interactions are common and play a vital role in optimization and discovery. By excluding them as part of

exploration, we greatly reduce the effectiveness of our experiments. Although still presented in some discipline-specific published literature, it is a poor strategy to gain understanding of processes or to optimize them.

We advocate for the use of the designed experiment approach, a mature but still growing field in statistics. Under this approach, a region of interest defined by the ranges of various inputs is studied, and, unlike the OFAT approach, the levels of more than one factor are intentionally changed for each run of the experiment. Methods based on this approach are more likely to identify the overall optimal operating conditions and establish a trustworthy causative relationship. Unlike the cautionary reminder that "correlation does not imply causation," with designed experiments, one actively seeks to establish causation. By manipulating variables deliberately, we move beyond mere associations to uncover true cause-and-effect relationships.

We advocate for the use of the designed experiment approach

In this approach, interaction effects can be discovered and characterized. W. Edwards Deming (1986, p. 89), who rarely mentioned designed experimentation explicitly, wrote the following:

> *Any step of the Shewhart (process improvement) cycle may need guidance of statistical methodology for economy, speed, and protection from faulty conclusions from failure to test and measure the effects of interactions.*

A primary purpose of experimentation is to obtain data to better understand the process of interest and guide improvements. In agriculture, for example, studying plant variety, fertilizer application, and irrigation levels can increase crop yield while controlling costs.

The use of experimentation in industrial applications became much more common in the U.S. in the 1980s. This was largely due to the influence of George E. P. Box, Genichi Taguchi, and Six Sigma

programs. In recent decades, the use of designed experiments has become much more widespread in industry and government, but we believe there are many more opportunities to expand their use and to design experiments more effectively.

"Big data" does not eliminate the need for experimentation, as discussed by Anderson-Cook and Lu (2023). Often big data has weaknesses, including having been obtained from observational studies (correlation does not imply causation), focusing on different aspects than the primary goal of a new study, or having gaps in the range of inputs that are of interest. Complementing big data with targeted designed experiments can greatly enhance the quality of results and improve their relevance to drawing meaningful and actionable conclusions.

"Big data" does not eliminate the need for experimentation

More recently, many organizations have found it beneficial to use computer experiments based on user-created mathematical models that mimic the physical process or product of interest. This approach can significantly reduce costs and accelerate experimentation. Additionally, it enables exploration of input combinations that may be difficult to explore in a physical setting.

The goal of this white paper is to demonstrate how designed experiments can be useful in the innovation process, both in improving existing products and processes, as well as in developing new products and processes. In a panel format, Jensen et al. (2012) discussed more generally how statistics, including experimentation, can facilitate innovation. There are several types of innovation, so we first discuss them and some of their features.

The nature of innovation

Breakthrough vs. incremental

Breakthrough innovation involves the development of a radical new product or idea, whereas incremental innovation involves the more modest improvement or adaptation of an existing process or product. To remain competitive and successful, companies must pursue both types of innovation. Montgomery and Silvestrini (2018) wrote, "It is our view that design of experiments is the most powerful statistical tool that is useful in enhancing both breakthrough and incremental innovation."

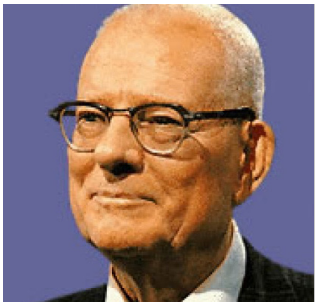


Figure 2: W. Edwards Deming

Simply put, organizations must innovate to adapt to changing environments and needs to survive. For a full discussion of strategy regarding innovation, which is beyond the scope of this white paper, we suggest Christensen and Raynor (2013).

Because of his continuing influence on those in the quality field, it is of interest to report what W. Edwards Deming (shown in Figure 2) said about innovation. Deming was primarily known for promoting quality improvement and increases in production efficiency, but he also emphasized the necessity of innovation.

> *One requirement for innovation is faith that there will be a future.*
>
> *Innovation, the foundation of the future, cannot thrive unless top management have declared unshakable commitment to quality and productivity. (Deming, 1986, p. 25)*
>
> *A good question for anyone in business to ask is, 'What business are we in?'*
>
> *...We must keep asking 'What product or service would help our customers more?' We must think about the future. What will we be making [five] years from now? [Ten] years from now? (Deming, 1993, p. 10)*

Preparation for the future includes ... constant scanning of the environment (technical, social, economic) to perceive need for innovation, new product, new service, of innovation of method. A company can to some extent govern its own future. (Deming, 1993, p. 55)

It is thus not sufficient to improve processes. There must also be constant improvement of design of product and service, along with the introduction of new product and service and new technology. All this is management's responsibility. (Deming, 1986, p. 135)

Deming taught that improving processes and products was necessary, but not sufficient. Innovation is also needed. This view is reflected in Figure 3, reproduced from Hockman and Jensen (2016).

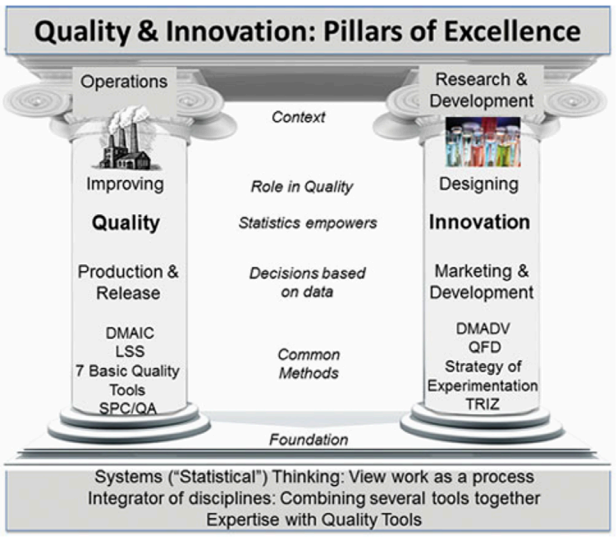


Figure 3: The Importance of Quality and Innovation, reproduced from Hockman and Jensen (2016).

Figure 3: The Importance of Quality and Innovation, reproduced from Hockman and Jensen (2016).

Creative destruction

Breakthrough and disruptive innovation can destroy the market of a current product, regardless of the product's quality or the efficiency of the production process. This ongoing evolution of the marketplace is known as creative destruction. Deming described how

buggy whips became obsolete when automobiles replaced horse-drawn carriages. More recent examples include personal computers replacing typewriters, floppy disks being replaced by CDs and USB flash drives, and numerous electronic devices being replaced by smart phones. Perhaps due to recognition of the ubiquitous phenomenon of innovation trumping quality, the Innovation Division of the American Society for Quality (ASQ) now has more than 5,500 members.

According to Deming (1993, p. 10), companies should think broadly about their business. He stated that it would have been better for manufacturers of automobile engine carburetors to define their business more generally as putting a stoichiometric mixture of fuel and air into the combustions chamber. By doing so, the manufacturers could have invented and developed something to do this better than carburetors, such as the fuel injection systems that eventually replaced them. Focusing on the functionality of the product allows for better responsiveness to opportunity than sticking with the current approach.

Anticipating what a customer might want is challenging business question

There are many implications for the increase of applications where creative destruction is prevalent. In a slowly evolving established market, organizations can prioritize incremental innovation. However, no improvements or enhancements to buggy whips would have protected them from becoming obsolete when disruption came. Understanding the trends of the market and customer needs becomes more important than focusing on direct competitors. Being the highest quality or best value typewriter manufacturer would offer no protection from the emergence of personal computers. Finally, anticipating what a customer might want as a future solution to a currently poorly defined or undefined problem is a substantially more challenging business question than responding to customer feedback for gradual improvements to an existing product.

Creative destruction can either be self-created or imposed by competition. Some companies, such as Google and Apple, include in

their strategic planning that they will intentionally work to disrupt and destroy their own products and processes with innovative replacements that redefine, update, or expand their customers' needs and wants. Companies that lack this forward-thinking approach are left to react to a changing marketplace without any way to shape how the new market evolves.

The role of the epiphany

Sometimes after a great deal of thought and work, an epiphany will occur where every detail of the solution to a problem falls into place. Herbert Robbins (1915-2001), the accomplished mathematician and statistician at Columbia University, said the following in an interview with Page (1984):

> *Over and over again I keep working at it, trying to understand something, which, after months or even years, turns out to be so simple that I should have seen it in the first ten minutes.*

In a similar vein, Drucker (1985) wrote the following:

> *All effective innovations are breathtakingly simple. Indeed, the greatest praise an innovation can receive is for people to say: "This is obvious. Why didn't I think of it?"*

A key point from these excerpts is that once an innovative idea or product has been discovered, it will often appear to be an obvious result. For this reason, we recommend that each person on the team involved in running an experiment predict in advance what the experimental results will show. This thoughtful anticipation of results, as well as the discussion that ensues when the experimental results are obtained and compared to predictions, can lead to improved understanding of the process.

Once an innovative idea or product has been discovered, it will often appear to be an obvious result.

Serendipity

Sometimes it pays to be lucky. The following products are among many that were discovered accidentally: the microwave oven, X-rays, penicillin, artificial sweeteners, Teflon, GORE-TEX fabric, and Post-it Notes. We recommend carefully examining any unusual or unanticipated experimental results and paying attention to outlying observations in the experimental data. Examining the potential of what has been observed in the light of other applications can lead to new products and processes. In drug discovery, for example, it is not uncommon that a "side effect" of a treatment for one health issue might suggest the use of the product for an entirely new purpose.

Along these lines, Box, Hunter, and Hunter (2005) wrote, "Discovering the unexpected is more important than confirming the known."

Serendipity thrives on the appreciation of unexpected discoveries, the thoughtful interpretation of results, and the willingness to embrace change.

We are continuously confronted with the need to adapt and evolve

Resistance to innovation

One should expect some resistance to any innovative idea, product, or process. In a 2001 interview with Daniel Peña, George Box stated the following:

> *..., I have found that initially original work is inevitably met with some hostility. For example, my first paper on response surfaces and my first paper in which the word 'robustness' appeared for the first time were both extremely difficult to get published. I think new ideas upset people.*

One should expect to encounter resistance when change is proposed and work to develop a consensus among the various stakeholders to be affected by an innovation. In our quickly evolving world, we are continuously confronted with the need to adapt and evolve to changing environments, markets, and rules.

Multiple discovery

Multiple discovery, also known as simultaneous invention, is the phenomenon that many scientific discoveries and inventions are made independently and nearly simultaneously by multiple scientists and inventors. Since being the first to market is a strong competitive advantage, it is important to keep the time required for product or process development to a minimum. Using designed experiments can be a key strategy to speed up the development process.

A compelling example of multiple discovery in the field of statistics is the survival curve estimation method simultaneously proposed by Edward L. Kaplan and Paul Meier to be used when some of the data values are right censored. This situation occurs in studies of medical interventions intended to extend human lifetimes. As discussed by Stalpers and Kaplan (2018), the two authors independently submitted papers containing an identical approach to the *Journal of the American Statistical Association*. The editor suggested that Kaplan and Meier work together to present a unified version of the methodology. The Kaplan and Meier (1958) paper was once ranked as the 11th most cited paper in all of science and currently has over 65,000 citations on Google Scholar.

The rate of multiple discovery seems to be increasing, perhaps due to the wider accessibility of information

Ogburn and Thomas (1922) documented 148 discoveries that were made independently by more than one scientist. These discoveries included the formulation of calculus and the discovery of oxygen. Many innovative inventions involve various kinds of simultaneous, overlapping, or disputed origins. Berkun (2010, p. 100) reported that this was true of the television, telephone, bicycle, motion pictures, MRI imaging, and automobile. The rate of multiple discovery in science, and likely in research and development organizations, seems to be increasing, perhaps due to the wider accessibility of information. Increasing the speed of innovation within a business can mean the difference between being first to market and not being in the conversation.

Application areas of designed experimentation

Types of designed experiments

An experiment is a test, or a series of tests, in which purposeful changes are made to input variables (called factors) of a process or system to observe and identify their impact on the output response(s) of interest. Understanding which factors affect the response(s) and the nature of the change is key to develop a model relating the response(s) to the levels of the factors. The model can then be used for improved understanding, process improvement, or other decision making. Sequential experimentation involves using the results of initial experiments to build on what has been learned to design follow-up experiments.

For free online introductions to designed experimentation, we recommend these: Design of Experiments (DOE) Course | JMP and the Coursera offering at Arizona State University, Design of Experiments Specialization [4 courses] (ASU) | Coursera.

Sometimes very simple experiments can be useful. Hockman and Jensen (2016) wrote, "We agree with the sentiment of Box (1990, pp. 367– 368):

> *'There are hundreds of thousands of engineers in this country, and even if the 23 factorial design was the only kind of [experimental] design they ever used, and even if the only method of analysis that was employed was to eyeball the data, this alone could have an enormous impact on the experimental efficiency, the rate of innovation, and the competitive position of this country.'"*

This statement describes the most rudimentary type of experiment and its analysis; many more options are available and refined

analyses can extract more knowledge from the data. The wide use of experiments (whatever their form) is highly desirable, and it is generally beneficial to design experiments carefully based on the specific goals of the experimentation, knowledge of the process, and the practical constraints.

Griffin et al. (2012) defined serial innovators as mature firms that are repeatedly successful in developing breakthrough innovations. They wrote, "Serial innovators use design of experiment methodologies to model total systems. When the product or process is too complex to model experimentally, they turn to system simulation methodologies to help them develop potential solutions."

It is beneficial to design experiments carefully based on the specific goals of the experimentation, knowledge of the process, and the practical constraints.

When thinking about which type of designed experiment to consider, it is important to start by identifying the objectives of the study:

- **Pilot experiments** are helpful for gaining the basic understanding about what the process is capable of and understanding if the accuracy and precision of measurement of the responses of interest are adequate for the goals of the study.
- **Exploratory or screening experiments** are often used to identify the most important factors impacting the response (as well as eliminate other factors that are not influential).
- **Response surface designs** (Myers et al., 2016) allow for modeling of the responses in functional form, so that the model can be used for understanding and identifying ideal regions where the process can generate promising response values.
- **Confirmation experiments** (Stevens and Anderson-Cook, 2019) are generally small experiments that allow for verification of the process settings that will generate the desirable outcomes.

For studies with multiple objectives, it is common to think in terms of a sequence of experiments, starting with a pilot study to understand

basics, moving to screening to identify important inputs, then building a model to codify the relationships and exploit them to find ideal regions in which to operate the process. A final step is to run a small confirmatory experiment to verify results before wide-scale implementation (Lu and Anderson-Cook, 2021).

It is important before beginning experimentation to think about which response(s) should be considered. Running a designed experiment requires investment and resources, so making sure to gather information about all attributes of interest can enhance learning and lead to greater understanding of the process or product. Experimenters can enhance their chance of a serendipitous innovation if they gather data on an array of responses to understand what is possible throughout the explored input space.

Experimenters can enhance their chance of a serendipitous innovation if they gather data on an array of responses

There are some other specialized types of experiments. Mixture experiments are used when each factor corresponds to a proportion of a given ingredient in a mixture. Thus, the factor levels must sum to 100%, adding a constraint to the design space. More recently, reliability experiments have been proposed where the goal is to increase the reliability of the product or process of interest (Rigdon et al., 2022). Censored lifetime data often result from such experiments. Some of these different types of designs will be revisited in the remainder of the paper.

Six Sigma applications

Designed experimentation is an important component in Six Sigma and similar types of programs, both in structured problem-solving approaches of define-measure-analyze-improve-control (DMAIC) and define-measure-analyze-design-verify (DMADV). The DMAIC approach is most useful for improving an existing product or process, whereas the DMADV approach is intended to be used when developing a new service, product, or process.

Montgomery and Woodall (2008) wrote, "Designed experiments are probably the most important statistical tool in the (DMAIC) Improve step. Designed experiments can be applied either to an actual physical process or to a computer simulation model of that process and can be used both for determining which factors influence the outcome of a process and for determining the optimal combination of factor settings." The intentional manipulation of the inputs allows causal relationships to be identified and leveraged.

In addition, Bisgaard (2006) wrote that Six Sigma provides a systematic approach to process and product improvement and innovation, and that design of experiments and computer experimentation are important tools within the product development process.

Figure 4, reproduced from Hockman and Jensen (2016), shows where designed experimentation has been useful throughout the DuPont product commercialization framework in assessing and optimizing candidate versions of the product or process under development.

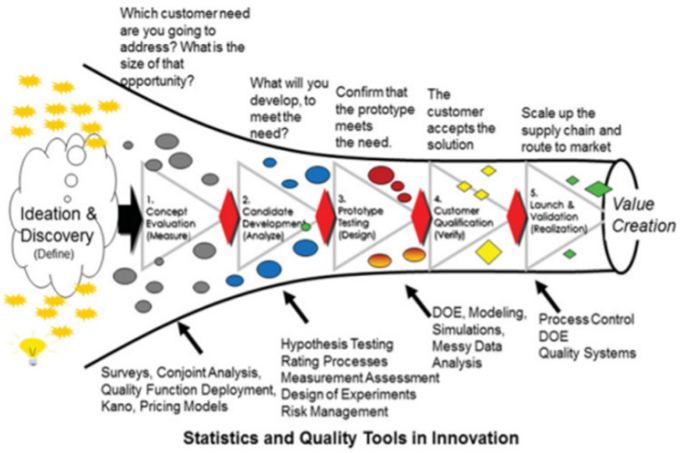


Figure 2. Product commercialization framework developed inside DuPont for commercializing new products. This is an example of a stage-gate methodology described in Cooper (2011).

Figure 4: The DuPont Commercialization Framework. From Hockman and Jensen (2016)

Computer experiments

As discussed by Thomke (2003, 2020), experiments based on computer models have been shown to be of considerable value in aiding innovation. While this approach can be much cheaper and quicker than building physical prototypes or experimenting with the actual physical system, it requires that an accurate model be created that approximates the true system well. Broadly speaking, there are two types of computer models.

1. Deterministic computer model. In this case, the output is completely deterministic, meaning that repeated runs at the same input factor settings result in the same value of the response. These types of models are widely used in engineering design applications and could involve finite element analysis, computational fluid dynamics, computational thermodynamics, or electrical circuit and device design software. These models often reflect current underlying understanding of a scientific or engineering mechanism. One design and analysis approach that is widely used for this is a space-filling designed experiment followed by fitting a response surface using a Gaussian process modeling approach. There are a variety of space-filling designs that exist (Santner et al., 2019), which can be easily adapted to different numbers and types of input factors, regions of interest, and varying emphases throughout the region (Lu et al., 2021). Sequential improvement of these models using Bayesian optimization has gained recent interest (Gramacy, 2020).

2. Stochastic simulation model. These are typically models of physical systems, such as manufacturing processes, service systems such as hospital emergency rooms, and other similar activities. These simulation models have inputs that are driven by random variables, such as the characteristics of jobs (such as processing times and due dates) and incoming units (such as patient arrival rates and presenting symptoms), so the outputs are random variables. These models are based on empirical patterns of processes and capture the

variability associated with human actions and behaviors. Experiments for these types of computer models typically involve the standard experimental designs that would be used with actual physical systems and processes.

Service process applications and online experimentation

Designed experiments are increasingly used in the nonmanufacturing world, such as in marketing, software testing, usability, choice of parameters within machine learning algorithms, and with online web-based experiments. See Antony et al. (2011) for a discussion of diverse examples of applications outside of manufacturing.

Thousands upon thousands of online experiments are being run to optimize webpage performance. Siroker and Kooman (2015), Thomke (2020), Haizler and Steinberg (2021), and Larsen et al. (2024) discussed A/B testing, which is frequently used for website improvement. This emerging area with new types of designed experiments exploits the high volume of users that can be involved in the testing. Websites can be adapted in real time to incoming information to allow continued learning while also seeking to minimize lost opportunity cost.

Quality by Design (QbD) in the pharmaceutical and biotech industries

Designed experimentation is a fundamental part of Quality by Design (QbD) in the pharmaceutical and biotech industries. It is a systematic approach to developing various treatments — and the manufacturing processes of those treatments — that begins with predefined objectives, emphasizes product and process understanding, and sets up process control based on sound science and quality risk management. It ultimately results in accelerated development and time to market.

Mixture experiments, response surface methods, and other types of experiments are an integral part of QbD. Response surface designs are used to determine regions of values of input variables that would lead to the production of acceptable drug products. See Politis et al. (2017), Fukuda et al. (2018) and Kenett et al. (2023, Chapter 6).

Gilman et al. (2021) discussed the use of designed experimentation in synthetic biology. Here, the emphasis is on the design and optimization of biological systems using engineering principles. These authors reviewed the literature and advocated the use of designed experimentation over the commonly used OFAT experiments that don't consider the inherent complexity of nature. Case studies involving designed experimentation in synthetic biology, as well as in many other applications, can be found at jmp.com.

Guidance on running experiments

Matching innovation opportunities to a designed experiment

Incremental innovation can potentially result from the use of both screening and response surface designs. Screening experiments may be incorporated to explore how manipulating some of the established input settings might alter the characteristics of the product or process. By considering alternatives to the status quo, there are often opportunities to improve either quality or efficiency. A good example was provided in the application discussed by Bullington, et al. (1992), in which the average lifetime of an industrial thermostat was increased by a factor of ten by a change in the raw material. Further, screening experiments permit exploration of inputs that can lead to serendipitous innovation.

Response surface designs can provide a way of characterizing the various responses with an estimated model, which allows for ideal combinations of the multiple responses to be simultaneously balanced and a suitable choice of inputs to be identified (Chapman et al., 2014).

Breakthrough innovation likely happens in the presence of considerably less background knowledge about what to expect for the product or process. In these cases, the entire sequence of pilot, exploratory, response surface, and confirmation experiments may be needed to gain a fundamental understanding, to explore which inputs are influential, and then to build sufficient understanding of the underlying relationships between potential inputs and critical characteristics of the new product or process. Broad exploration of the ranges of diverse inputs is important for identifying and understanding regions where a new viable solution is attainable.

Breakthrough innovation happens in the presence of considerably less background knowledge about what to expect for the product or process.

When a serendipitous or epiphanous moment occurs, design of experiments can help to fill in gaps of knowledge and develop models to characterize quality properties. Sometimes these unplanned moments hinge on changing the focus or utility of the product to serve a different use or advantage for the customer.

Choosing the right design

Choosing the right experimental design begins with defining clear goals. Defining success for a specific study — innovation, efficiency, or specific outcomes — may require distinct types of experimental data.

Choosing the right experimental design begins with defining clear goals

Once the goal(s) of the experiment are established, one must consider the logistical aspects of the experiment. Which factors to include involves trade-offs: too many increase the size of the experiment, while excluding potentially important factors does not allow assessment of their impact. Factor ranges balance between being too narrow to potentially miss regions of desirable performance versus being too wide that operational viability is challenged. More complicated anticipated relationships require additional levels of each factor to be included in the design, which adds to the overall size of the experiment needed. If some regions of the input space are nonviable, then constraints can be included to tailor the experimental region to ensure resources are not wasted.

The JMP white paper (Anderson-Cook, 2022) and the Statistically Speaking presentation at Design of Experiment's Crucial Step 0: Choosing the Right DOE Option - (2023-US-... - JMP User Community provide some guidance on how to construct and choose between a collection of designs of different sizes — and with different strengths and weaknesses — using design assessment tools that quantify appropriate metrics to match the objectives of the study.

Tips for successful implementation

Create the team that will plan and execute the experiment. The first step to a successful experiment is forming a team of representatives from all parties that have an interest in the motivating problem and solution. The typical stakeholders are engineering, quality assurance, manufacturing, marketing, management, customers, and the operating personnel who will be involved in the hands-on running of the experiment (who usually have much insight and, sadly, are often ignored). This team approach, which leverages cognitive and background diversity, is essential to success.

Develop all ideas about the problem and the objectives of the experiment. A crucial mistake that we have seen all too often is to let a single "expert" design the experiment. The expert will often design an experiment that is much too narrow in scope, does not include the right factors, and misses important opportunities to explore and learn. The resulting failure can discourage future use of designed experiments in the organization.

Create a process with clearly defined steps. Planning and conducting a designed experiment is a process. Montgomery (2020) gives a seven-step process, which we briefly discuss below. The papers by Coleman and Montgomery (1993) and Freeman et al. (2013) also contain useful information on planning experiments.

STEP 1 **RECOGNITION OF AND STATEMENT OF THE PROBLEM**

- **Prepare a list of specific problems or questions that are to be addressed by the experiment.** It is helpful for the team to have a clear statement of the problem to help them better understand the phenomenon being studied and the ultimate solution to the problem.

- **Keep the overall objectives of the experiment in mind.** There are several broad reasons for running experiments and each type of experiment will generate its own list of specific questions that need to be addressed. Some (but by no means all) of the objectives for running experiments are included in the Appendix.

- **Experiment sequentially and invest no more than 25 percent of available resources in the first experiment.** It is usually a major mistake to design a single, large, comprehensive experiment at the start of a study. A successful experiment requires knowledge of the important factors, the ranges over which these factors should be varied, the appropriate number of levels to use, and the proper units of measurement for these variables. Generally, at the outset we do not know this information, but we learn more as we go along.

As an experimental program progresses, we often drop some input variables, add others, change the region of exploration for some factors, or add new response variables. Consequently, we usually experiment sequentially and, generally, no more than about 25 percent of the available resources should be invested in the first experiment. This rule of thumb ensures that sufficient resources are available to respond to what is learned, adjust to new information, and perform confirmation runs. This strategy will ultimately help to accomplish the final objective of the experiment.

SELECTION OF THE RESPONSE VARIABLES

- **Be certain that response variables provide useful information about the system and are measurable.** Most often, the average or standard deviation (or both) of the measured characteristics will be included in the set of response variables.

- **Justify investment with multiple responses, which are common and often beneficial.** Setting up and executing experiments require considerable resources, and collecting data on multiple attributes

helps to justify the investment. The experimenters must decide how each response will be measured, and address issues such as how any measurement system will be calibrated and how this calibration will be maintained during the experiment.

The gauge or measurement system capability (or measurement error) is also an important factor. If the quality of measurements for the response is inadequate, only relatively large factor effects will be detected by the experiment or additional replication will be required. In some situations where gauge capability is poor, the experimenters may decide to measure each experimental unit several times and use the average of the repeated measurements as the observed response.

- **Identify issues related to defining the responses of interest and how they are to be measured before conducting the experiment.** Some designed experiments are employed to specifically study and improve the performance of measurement systems (Burdick et al., 2005).

STEP 3 CHOICE OF FACTORS, THEIR LEVELS, AND THEIR RANGES

- **Factors can be classified as either potential design factors or nuisance factors.** When considering the factors that may influence the performance of a process or system, the experimenters usually discover that these factors can be classified as either potential design factors or nuisance factors. The potential design factors are those factors that the experimenters may wish to vary in the experiment.

Some useful classifications are design factors, held-constant factors, and allowed-to-vary factors. The design factors are the factors selected for study in the experiment. Held-constant factors are variables that may exert some effect on the response, but for purposes of the present experiment, these factors are not of interest, so they will be held at a specific level. We often assume that the effects of held-constant factors and allowed-to-vary factors are relatively small.

- **The blocking principle is often useful in dealing with controllable nuisance factors.** Nuisance factors, on the other hand, may have large effects on the responses that must be accounted for, yet we may not be interested in them in the context of the present experiment. Nuisance factors are often classified as controllable, uncontrollable, or noise factors. A controllable nuisance factor is one whose levels may be set by the experimenter.

  For example, the experimenter can select different batches of raw material or different days of the week when conducting the experiment. The blocking principle is often useful in dealing with controllable nuisance factors. If a nuisance factor is uncontrollable in the experiment but can be measured, analysis of covariance can often be used to compensate for its effect. When a factor that varies naturally and uncontrollably in the process can be controlled for purposes of an experiment, we often call it a noise factor. In such situations, our objective is usually to find the settings of the controllable design factors that minimize the variability transmitted from the noise factors. This is sometimes called a process robustness study or a robust parameter design problem.

- **Investigate all factors that may be of importance.** Once the experimenters have selected the design factors, they must choose the ranges over which these factors will be varied and the specific levels that will be used. Thought must also be given to how these factors are to be controlled at the desired values and how they are to be measured. It is crucial to investigate all factors that may be of importance and not to be overly influenced by past experience, particularly when we are in the early stages of experimentation or when the process is not very mature. Small pilot runs may be useful at this stage of planning.

When the objective of the experiment is factor screening or process characterization, it is usually best to keep the number of factor levels low. Generally, two levels work very well in factor screening studies. Recent work provides options so that continuous factors have a few runs at a central value, allowing detection of large quadratic effects.

Choosing the experimental region of interest is also important. In factor screening, the region of interest should be relatively large — that is, the range over which the factors are varied should be broad. As we learn more about which variables are important and which levels produce the best results, the region of interest in subsequent experiments will usually become narrower.

STEP 4 CHOICE OF EXPERIMENTAL DESIGN

- **When choosing an appropriate design, consider the number of runs, select a suitable run order for the experimental trials, and then determine if blocking or other randomization restrictions are involved.** It is also important to keep the experimental objectives in mind. Interactive statistical software packages (such as our favorite, JMP) support this phase of experimental design. Experimenters can enter information about the goal of the experiment, number of factors, levels, and ranges, and then be guided to a selection of designs for consideration or receive a recommendation of a particular design. JMP software also provides useful diagnostic information about how each design will perform across a variety of relevant design performance metrics.

- **Design selection also involves thinking about and selecting a tentative empirical model to describe the results.** The model is just a quantitative relationship (equation) that connects the response and the important design factors. In many cases, a low-order polynomial model will be appropriate. First-order models are used extensively in screening or characterization experiments. Second-order models are often used in optimization experiments.

STEP 5 PERFORMING THE EXPERIMENT

- **When running the experiment, it is vital to monitor the process carefully to ensure that everything is being done according to plan.** Errors in experimental procedure at this stage can weaken or destroy experimental validity. One of the most common mistakes that we have encountered is that those conducting the experiment failed to set the variables to the proper levels on some runs. Another common mistake is to break the randomization of the runs by reordering experiments out of convenience.

- **Someone should be assigned to check factor settings before each run.** Sticking with the plan and implementing it as intended can protect against complications that are challenging or impossible to unravel without additional resources being spent. Up-front planning to prevent this type of mistake is crucial to success. It is easy to underestimate the logistical and planning aspects of running a designed experiment in a complex manufacturing or research and development environment.

- **A few trial or pilot runs are often helpful prior to conducting the experiment.** (Coleman and Montgomery, 1993). These runs provide information about consistency of experimental material, a check on the measurement system, a rough idea of experimental error, and a chance to practice the overall experimental technique. This also provides an opportunity to revisit the decisions made in Steps 1–4, if necessary.

STEP 6 STATISTICAL ANALYSIS OF THE DATA

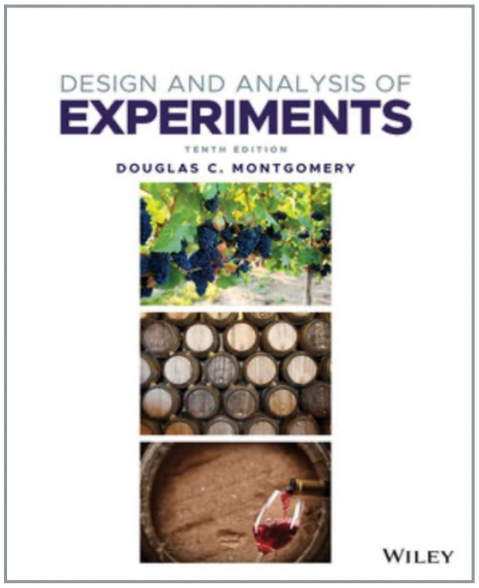
- **Software provides a seamless, direct interface to the statistical analysis, and we often find that simple graphical methods play an important role in data analysis and interpretation.** Hypothesis testing and confidence interval estimation procedures are very useful in analyzing data from a designed experiment. It is also usually very helpful to present the results of many experiments in terms of an

empirical model, i.e., an equation derived from the data that expresses the relationship between the response and the important design factors. Residual analysis and model adequacy checking are also important analysis techniques. The primary advantage of statistical methods is that they add objectivity to the decision-making process. Statistical techniques coupled with good engineering or process knowledge, and common sense will usually lead to sound conclusions.

STEP 7 CONCLUSIONS AND RECOMMENDATIONS

- **Validate conclusions and connect results to organization goals.**

Once the data have been analyzed, the experimenter must draw practical conclusions about the results and recommend a course of action. Follow-up runs and confirmation testing should also be performed to validate the conclusions from the experiment and ensure that what has been learned can be implemented. Graphical summaries are often useful in this stage, particularly in highlighting and communicating the results to others. When sharing results, it is critical to connect them to the goals and missions of the organization.



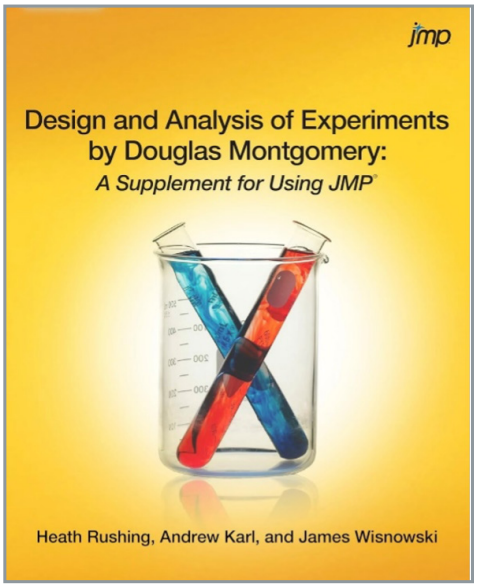


Figure 5: Recommended books on designed experimentation

Conclusion

Ongoing innovation is required for ongoing success. Having high-quality products and efficient production processes is necessary, but not sufficient. Designed experimentation plays an important role in the innovation process. We have provided some guidelines on how to conduct effective experiments.

It is important to acknowledge that all experiments are designed experiments. The critical issue is whether they are well-designed or not. Good pre-experimental planning will usually lead to a successful experiment.

There are many books and resources on designed experimentation. We recommend Montgomery (2020) and Rushing et al. (2014).

References

Anderson-Cook, C. (2022). Choosing the Right Design – with an Assist from JMP's Design Explorer, JMP white paper. Choosing the Right Design – with an Assist from JMP's Design Explorer | JMP

Anderson-Cook, C.M., Alley, M. M., Roygard, J.K.F., Khosla, R., Noble, R. B., & Doolittle, J. A. (2002). Differentiating Soil Types Using Electromagnetic Induction and Crop Yield Maps. *Soil Sciences Society of America Journal*, 66, 1562-1570.

Anderson-Cook, C. M., Goldfarb, H., Borror, C., Montgomery, D. C., Canter, K. G., & Twist, J.A. (2004). Mixture and Mixture-Process Variables Experiments for Pharmaceutical Applications. *Pharmaceutical Statistics*, 3, 247-260.

Anderson-Cook, C. M., & Lu, L. (2023). Is designed data collection still relevant in the big data era? *Quality and Reliability Engineering International*, 39(4), 1085-1101.

Antony, J., Coleman, S., Montgomery, D. C., Anderson, M. J., & Silvestrini, R. T. (2011). Design of experiments for non-manufacturing processes: benefits, challenges and some examples. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 225(11), 2078-2087.

Berkun, S. (2010). *The Myths of Innovation. Sebastopol*, CA: O'Reilly Media, Inc.

Bisgaard, S., & De Mast, J. (2006). After Six Sigma—What's next? *Quality Progress*, 39(1), 30–36

Bisgaard, S. (2012). The future of quality technology: From a manufacturing to a knowledge economy & from defects to innovations. *Quality Engineering*, 24(1), 30-36.

Box, G. E. P. (1990). George's column. *Quality Engineering*, *2*, 365–369.

Box, G. E. P. & Woodall, W. H. (2012). Innovation, quality engineering, and statistics. *Quality Engineering*, 24(1), 20-29.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). Statistics for Experimenters: Design, Innovation, and Discovery. In Wiley series in probability and statistics. Hoboken, NJ: Wiley.

Bullington, R. G., Lovin, S., Miller, D. M., & Woodall, W. H. (1993). Improvement of an industrial thermostat using designed experiments. *Journal of Quality Technology*, 25(4), 262-270.

Burdick R. K., Borror C. M., Montgomery, D.C. (2005). *Design and Analysis of Gauge R&R studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

Chapman, J., Lu, L., & Anderson-Cook, C.M. (2014). Process optimization for multiple responses utilizing the Pareto front approach. *Quality Engineering* 26, 253–268.

Christensen, C. M. (2013). The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Harvard Business Review Press.

Christensen, C., & Raynor, M. (2013). *The Innovator's Solution: Creating and Sustaining Successful Growth*. Harvard Business Review Press.

Coleman, D. E., & Montgomery, D. C. (1993). A systematic approach to planning for a designed industrial experiment. *Technometrics*, 35(1), 1-12.

Cooper, R. G. (2011). *Winning at New Products: Creating Value through Innovation*. 4th ed. New York, NY: Basic Books.

Deming, W. E. (1986). *Out of the Crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, Mass.

Deming, W. E. (1993). *The New Economics for Industry, Government, and Education*. Massachusetts Institute of Technology, Center for Advanced Engineering Study, Cambridge, Mass

Drucker, P. F. (1986). *Innovation and Entrepreneurship: Practice and Principles*. Harper & Row.

Freeman, L.J., Hutto, G., & Mackertich, N., (2014). Scientific Methods for Improving DoD Test and Evaluation. *ITEA Journal of Test and Evaluation* 35(1), 31.

Freeman, L.J., Ryan, A.G., Kensler, J. L. K., Dickinson, R.M., & Vining, G.G. (2013). A tutorial on the planning of experiments. *Quality Engineering*. 25, 315-332.

Fukuda, I. M., Pinto, C. F. F., Moreira, C. D. S., Saviano, A. M., & Lourenço, F. R. (2018). Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). *Brazilian Journal of Pharmaceutical Sciences*, 54.

Garud, S. S., Karimi, I. A., & Kraft, M. (2017). Design of computer experiments: A review. *Computers & Chemical Engineering*, 106, 71-95.

Ghawaly, J. M., Nicholson, A., Peplow, D. E., Anderson-Cook, C. M., Myers, K. L., Archer, D. E., Willis, M. J., & Quiter, B. J. (2020). Data for training and testing radiation detection algorithms in an urban environment. *Scientific Data* 7, 328.

Gilman, J., Walls, L., Bandiera, L., & Menolascina, F. (2021). Statistical design of experiments for synthetic biology. *ACS Synthetic Biology*, 10(1), 1-18.

Gramacy, R.B., (2020). Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences. Chapman and Hall/CRC.

Griffin, A., Price, R. L., & Vojak, B. (2012). *Serial Innovators: How Individuals Create and Deliver Breakthrough Innovations in Mature Firms*. Stanford University Press.

Haizler, T. & Steinberg, D. M. (2021). Factorial designs for online experiments, *Technometrics*, 63(1), 1-12.

Hashiba, A., Toyooka, M., Sato, Y., Maeki, M., Tokeshi, M., & Harashima, H. (2020). The use of design of experiments with multiple responses to determine optimal formulations for in vivo hepatic mRNA delivery. *Journal of Controlled Release*, 327, 467-476.

Hockman, K. K., & Jensen, W. A. (2016). Statisticians as innovation leaders. *Quality Engineering*, 28(2), 165-174.

Jensen, W., Anderson-Cook, C., Costello, J.A., Doganaksoy, N., Hoerl, R.W., Janis, S., O'Neill, J., Rodebaugh, B. & Snee, R.D. (2012). Statistics to facilitate innovation: a panel discussion. *Quality Engineering*, 24(1), 2-19.

JMP Webinar Series: Smarter Innovation with Design of Experiments | JMP

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457-481.

Kenett, R. S., Zacks, S., & Gedeck, P. (2023). *Industrial Statistics: A Computer-based Approach with Python*. Springer Nature.

Larsen, N., Stallrich, J., Sengupta, S., Deng, A., Kohavi, R., & Stevens, N. T. (2024). Statistical challenges in online controlled experiments: A review of A/B testing methodology. *The American Statistician*, 78(2), 135-149.

Lekivetz, R., Morgan, J. (2021). On the Testing of Statistical Software. *Journal of Statistical Theory and Practice* 15(4), article number 76.

Lu, L., Anderson-Cook, C.M. (2021). Strategies for sequential design of experiments and augmentation. *Quality and Reliability Engineering International* 37(5) 1740-1757.

Lu, L., Anderson-Cook, C.M., Ahmed, T. (2021). Non-uniform space-filling designs. *Journal of Quality Technology* 53(3) 309-330.

Lujan-Moreno, G. A., Howard, P. R., Rojas, O. G., Montgomery, D. C. (2018). Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. *Expert System with Applications*, 109, 195-205.

Montgomery, D. C. (1999). Experimental design for product and process design and development. *Journal of the Royal Statistical Society Series D: The Statistician*, 48(2), 159-177.

Montgomery, D. C. (2020). *Design and Analysis of Experiments*, 10th edition, John Wiley & Sons, Inc.

Montgomery, D. C., & Silvestrini, R. T. (2018). Design of experiments: A key to successful innovation. *In Frontiers in Statistical Quality Control* 12 (pp. 279-291). Springer International Publishing.

Montgomery, D. C., & Woodall, W. H. (2008). An overview of six sigma. *International Statistical Review/Revue Internationale de Statistique*, 329-346.

Morgan, J.C., Omell, B., Matuszewski, M., Miller, D. C., Shah, M. I., Benquet, C., Knarvik, A.B.N., de Cazenove, T., Anderson-Cook, C. M., Ahmed, T., Tong, C., Ng, B., & Bhattacharyya, D. (2021). Application of sequential design of experiments (SDoE) to large pilot-scaled solvent-based CO2 capture process at Technology Centre Mongstad (TCM). Proceedings of 15th International Conference on Greenhouse Gas Control Technology, GHGT-15.

Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M. (2016). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 4th Ed. New York: Wiley.

Ogburn, W. F., & Thomas, D. (1922). Are inventions inevitable? A note on social evolution. *Political Science Quarterly*, 37(1), 83-98.

Page, W. (1984). An interview with Herbert Robbins. *The Two-year College Mathematics Journal* 15(1), 2-24.

Peña, D. (2001). George Box: An interview with the International Journal of Forecasting. *International Journal of Forecasting* 17, 1-9.

Politis, S.N., Colombo, P., Colombo, G., & Rekkas, D. M. (2017). Design of experiments (DoE) in pharmaceutical development. *Drug Development and Industrial Pharmacy*, 43(6), 889-901.

Rigdon, S. E., Pan, R., Montgomery, D. C., & Freeman, L. (2022). *Design of Experiments for Reliability Achievement*. John Wiley & Sons.

Román-Ramírez, L. A., & Marco, J. (2022). Design of experiments applied to lithium-ion batteries: A literature review. *Applied Energy*, 320, 119305.

Rushing, H., Karl, A., & Wisnowski, J. (2014). *Design and Analysis of Experiments by Douglas Montgomery: a supplement for using JMP*. SAS Institute.

Santner, T. J., Williams, B. J., Notz, W. I., & Williams, B. J. (2018). *The Design and Analysis of Computer Experiments*. 2nd Ed. New York: Springer.

Simpson, J.R., Listak, C.M., & Hutton, G.T. (2013). Guidelines for planning and evidence for assessing a well-designed experiment. *Quality Engineering*, 25 333–355.

Siroker, D., & Koomen, P. (2015). *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*. John Wiley & Sons, Inc.

Stalpers, L. J., & Kaplan, E. L. (2018). Edward L. Kaplan and the Kaplan-Meier survival curve. *BSHM Bulletin: Journal of the British Society for the History of Mathematics*, 33(2), 109-135.

Stevens, N.T., Anderson-Cook, C.M. (2019). Design and analysis of confirmation experiments. *Journal of Quality Technology* 51(2), 109-124.

Summers, A. J., Devadhasan J. P., Gu, J., Montgomery, D. C., Fischer, B., Gates-Hollingsworth, M. A., Pflughoeft, K. J., Vo-Dinh, T., AuCoin, D. P., & Zenhausern, F. (2022). Optimization of antibody microarray printing process using a designed experiment. *ACS Omega*, 7(36), 32262-32271.

Thomke, S. H. (2003). *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*. Harvard Business Press.

Thomke, S. H. (2020). *Experimentation Works: The Surprising Power of Business Experiments*. Harvard Business Press.

Yuangyai, C., & Nembhard, H. B. (2015). Design of experiments: a key to innovation in nanotechnology. In *Emerging Nanotechnologies for Manufacturing* (pp. 230-254). William Andrew Publishing.

Appendix: Common objectives for designed experiments

Factor screening or characterization
When a system or process is new, it is usually important to learn which factors have the most influence on the response(s) of interest. Often there are many factors suggested, which usually indicates that the experimenters do not know much about the system. As a result, screening is essential if we are to efficiently get the desired performance from the system. Screening experiments are extremely important when working with new systems or technologies so that valuable resources will not be wasted using best guess or OFAT approaches.

Optimization
After the system has been characterized and we are reasonably certain that the important factors have been identified, the next objective is usually optimization. That is, find the settings or levels of the important factors that result in desirable values of the response. For example, if a screening experiment on a chemical process results in the identification of time and temperature as the two most important factors, the optimization experiment may have as its objective finding the levels of time and temperature that maximize yield. An optimization experiment is usually a follow-up to a screening experiment. It would be very unusual for a screening experiment to produce the optimal settings of the important factors.

Confirmation
In a confirmation experiment, the experimenter is usually trying to verify that the system operates or behaves in a manner that is consistent with some theory, past experience, or prior experimentation. Moving a new manufacturing process to full-scale production based on results found during experimentation at a pilot plant is a well-suited situation for confirmation experiments – that is, are the same factors and settings that were determined during development work appropriate for the full-scale process?

Discovery
In discovery experiments, the experimenters are usually trying to determine what happens when we explore new materials, or new factors, or new ranges for factors. Discovery experiments often involve screening several (perhaps many) factors. In the pharmaceutical industry, scientists are constantly conducting discovery experiments to find new materials or combinations of materials that will be effective in treating disease.

Robustness
These experiments often address such questions as, under what conditions do the response variables of interest seriously degrade? Or what conditions would lead to unacceptable variability in the response variables? A variation of this type of design is determining how to set the factor levels in the system that can be controlled to minimize the variability transmitted into the response from factors that cannot be controlled very well.

Reliability and system life testing
Experiments to evaluate product reliability can be variations of the experiments to evaluate robustness described above. Often, they are experiments conducted to determine the product failure rate under typical conditions of use. Life testing experiments at conditions of use could require a very long time to complete, so many of these experiments use accelerating factors to intentionally increase the failure rate. Typical acceleration factors include temperature, humidity, and vibration. Units are tested at extremes of these factors to induce failures faster, then inference is drawn about failure rates at conditions of use by extrapolating from an appropriate statistical model such as one based on the Arrhenius equation.