Design Of Experiments

Improve Quality and you automatically improve productivity
– W. Edwards Deming

		Factors			
		Factor A	Factor B	Factor C	Factor D
Treatments	1	+	+	+	+
	3	+	-	-	+
	5	-	+	-	+
	7	-	-	+	+
	9	+	+	+	-
	11	+	-	-	-
	13	-	+	-	-
	15	-	-	+	-

A **designed experiment** is one of the best tools to **improve quality and productivity** at the same time!

A **design of experiments** or **DOE** is a statistical method that allows you to **study and quantify the relationship** between the **inputs** (factors) and **outputs** (responses) of a **process or product**.

The DOE tool is powerful in its ability to **study multiple factors (inputs) simultaneously** to determine their **effect on the response (outputs)**.

Used properly, a DOE can **optimize processes, improve quality, lower costs and improve your operations.**

DOE is commonly used during **Product/Process Design** and **Continuous Improvement**.

This chapter is laid out into **6 sections** leading you through the must know **DOE topics, concepts and techniques.**

Section 1 is the **basic terminology** used within the world of **DOE** which include **Factors (Inputs), Response (Outputs), Levels, Treatments, Error, Replication and Robustness**.

Section 2 is the basic process of how to **plan, organize, execute and analyze a well-designed experiment**.

The goal of this section is to help you identify the proper design to use for your experiment and the goal you're trying to accomplish.

Section 3 are the **critical design principles** that must be applied to a designed experiment which include **blocking, replication, sample size, power, efficiency, interactions and confounding**.

Section 4 is an introduction to the simplest of DOE's which is the **One Factor Experiment**. Within this section we will also refresh ourselves with **ANOVA**, which is the most common analysis technique that is paired up with a DOE.

Section 5 is the more complex **Full Factorial DOE** with an example**.**

Section 6 is the **Fractional Factorial DOE**.

Out of Scope for this Chapter

This chapter is focused on the **core concepts** and **common designs** (full factorial, fractional factorial, etc) within DOE.

I've excluded many of the more complex designs that can be used, as these are out of scope of the Six Sigma Green Belt Certification. These include Plackett-Burman Designs, Orthogonal Arrays and Response surface Designs (Central Composite Designs, Box-Behnken Designs).

Lastly, this chapter is focused on **two-level designs**. Designs of three levels, or mixed-level designs are out of scope.

DOE Terminology

Before we jump into the **concepts and techniques of DOE**, it's important to align on the **terminology**.

All of this terminology is centered around the **idea of a process**.

The Process Is Where It Starts

Every process is the same, and every process is different.

The process model is central to the idea of DOE.

Every process has 3 common features: **inputs**, the **process** and **outputs**.

The **inputs** are also commonly referred to as **Factors** or **Independent Variables**.

The **outputs** are also commonly called **Responses** or **Dependent Variables**.

The **process** is the **how we transform inputs into outputs**.

Let's use the classic example of **baking a cake** to demonstrate **how a process works**, and then how we would **design an experiment** to bake the most delicious cake ever.

Let's **start with the end in mind** and talk about **outputs** which are also called **response variables** or **dependent variables**.

Dependent Variables (Response)

Outputs (**response variables**) represent the **outcome** of a process or experiment.

These **dependent variables (Responses)** can be **quality attributes**, **reliability attributes**, **dimensional/functional requirements**, material requirements or **continuous improvement** metrics (yield, capacity, cycle time, etc.).

In the cake example, our **major output** of the process is **taste** – the cake should be delicious.

There are **other response outputs to consider** and include **how the cake looks visually**, how much the **cake costs**, the **time** required to cook the cake, and the **size** of the final cake.

Let's jump to the **other side of the process** now and cover **inputs**.

Independent Variables (Factors)

When we say **independent variables (x)** we are talking about the **inputs** or **factors** associated with your process.

These inputs can be **controllable** or **uncontrollable (noise)**.

In the cake example, **controllable inputs** include the raw materials (ingredients), the supplier of each ingredient, and the **process inputs** such as the temperature that we bake the cake at, or the baking time.

Controllable inputs can be modified within the experiment or process.

There are also **factors** associated with your process that are **uncontrollable (noise)** which can also have a major impact on your outputs.

Oftentimes, the purpose of an experiment is to reduce the impact that the **uncontrollable factors** have on our output – this result would mean that **process has been made robust to the uncontrollable factors**. More on this later.

Oftentimes **uncontrollable factors** are impossible to control in actual production, but can sometimes be controlled during an experiment to study their impact.

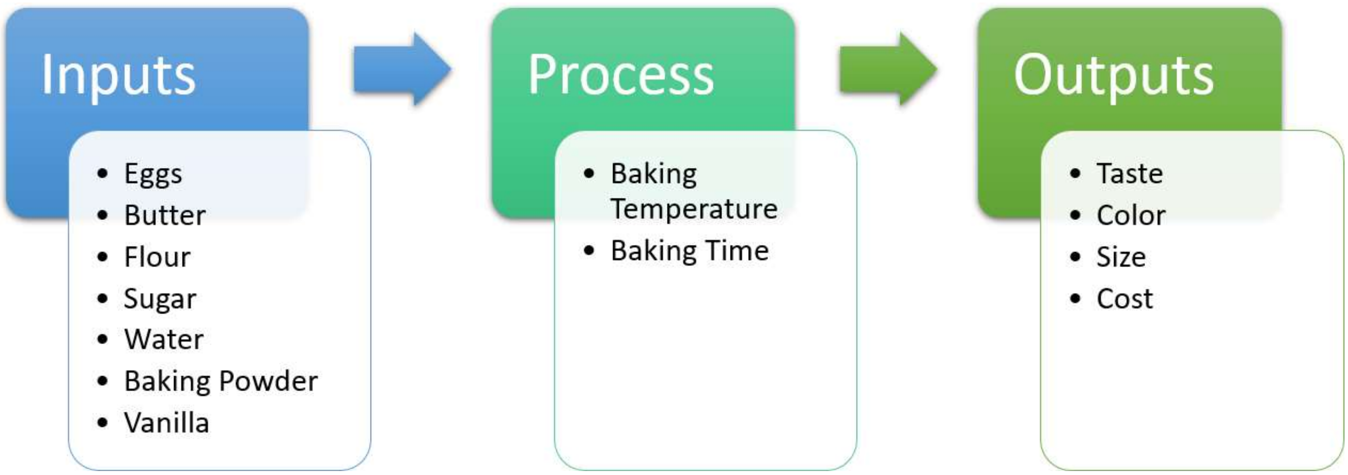
In the example of baking a cake there are factors that will influence the outputs but that are uncontrollable.

These **uncontrollable factors** include the altitude of the person baking the cake, or the ambient humidity of the environment of the person, or the type of oven being used (Convection/Conventional or Gas/Electrical, etc) or the location of the cake within the oven.

These **uncontrollable factors** are sometimes cause **nuisance factors** or **noise factors** because they can cause problems.

Later on, we will discuss how we can use techniques like **blocking and randomization** to minimize the variation created by these factors during an experiment.

Let's look at what our cake process looks like now with the inputs, process and outputs more clearly defined:



Noticed I've excluded the uncontrollable factors and only listed the inputs, but it's important to understand the uncontrollable factors so that you can design a robust process.

Levels

A level refers to specific settings of a factor.

For example, in the cake experiment, we could have 2 levels for the Egg Factor with one being a 2-egg cake, and the other being a 3-egg cake.

Or we could have 2 levels for the baking temperature, the high level at 400 degrees and the low level at 300 degrees.

For the purposes of this chapter, we will be focused on **2-level experiments**, where a **high and low level** will be defined for each factor which are commonly shown as **+ (high) and – (low)**.

We will not be covering experiments with 3 or more levels.

Treatment

A treatment is a unique combination of factors and levels within an experiment.

Let's say we wanted to run a full factorial experiment that only analyzed **4 factors** (baking temperature, baking time, eggs and sugar) at **2 levels** each.

	Temperature	Time	Eggs	Sugar
Treatment 1	+	+	+	+
Treatment 2	+	+	+	-
Treatment 3	+	+	-	+
Treatment 4	+	+	-	-
Treatment 5	+	-	+	+
Treatment 6	+	-	+	-
Treatment 7	+	-	-	+
Treatment 8	+	-	-	-
Treatment 9	-	+	+	+
Treatment 10	-	+	+	-
Treatment 11	-	+	-	+
Treatment 12	-	+	-	-
Treatment 13	-	-	+	+
Treatment 14	-	-	+	-
Treatment 15	-	-	-	+
Treatment 16	-	-	-	-

We can define a **"high" and "low" level** for each **factor**.

- **Temperature levels** might be 400° F and 350° F.
- **Time levels** might be 20 minutes and 15 minutes.
- **Eggs levels** might be 3 eggs and 2 eggs.
- **Sugar levels** might be 2 cups and 1 cup.

This **experiment** would have **16 unique treatments** associated with it, all with a unique combination of levels for each factor.

For example, **Treatment 8** would be the unique combination of 400° F baking temperature, 15 minutes baking time, 2 eggs and 1 cup of sugar.

Before we move on though, let's now talk about Error.

Random Error

Random error is the variation in your experimental results caused by both **controllable** and **uncontrollable factors (noise)** or **simply the random variation** in the response variable.

In SPC, when we talk about normal, random, inherent variation, this is similar to the idea of random error. It is the expected, normal, random variation associated with your response variable.

Blocking, replication and randomization are three tools that can be used to **reduce or eliminate the random error** which we will go over below.

Note – an **assumption within ANOVA** is that this **random error is normally distributed** with a mean of zero. Confirmation of this assumption during the analysis phase of a DOE is often required.

Systematic Error

There is another type of error that is **systematic in nature**, and is not related to the natural, random, inherent variation in your response variable. This error is not random in nature and affect all of your measurements in some way.

The classic example of systematic nature is **measurement bias** or **measurement error**. Human bias in the experiment can also be an example of **systematic bias**.

Let's say you execute an experiment and use a gage that it out of calibration. This unstable measurement system can introduce significant variation in your response variable.

Systemic error also occurs if your process is **not stable or in control**. If your process is under the influence of special cause variation, then it may have higher than expected levels of variation in the response variable.

Caution should be taken to eliminate systematic errors so that only the natural variation remains, because systemic errors can absolutely destroy the accuracy and precision of your experimental results.

Experimental Error

If you were to run your DOE 10 times, you'd like get 10 different sets of results. Now hopefully, if you've done your job correctly, and you've eliminated systematic error, and reduced random error, then those results would be fairly similar.

And this is the idea of experimental error which is the variation in the response variable of virtually identical test conditions (replicates).

If this error is too large, it has the power to wreck your experiment, leaving you conclusion-less.

Reducing **experimental error** increases the **accuracy of your conclusions** about the **effect of each factor**.

Replication

To minimize the experimental error, **replication** can be used.

Replication is the act of performing an experiment all over again – from start to finish, not simply remeasuring the response variable.

Each repetition of an experiment is called a **replicate**.

The estimate of the effects of each factor within an experiment becomes more precise when we replicate an experiment.

The first result of an experiment could be due to luck or chance or random variation in the response variable. As we replicate a result multiple times, our **results become more precise.**

Replicating an experiment gives confidence that **a result is repeatable** and not simply the result of random variation. Replication also allows for enough samples and **degrees of freedom to study interaction effects**, etc.

Replication also helps you **better estimate the random error** associated with your process. This helps during the **ANOVA** analysis phase.

Replicating an experiment increases the sample size and results in a more precise conclusion.

Design Principles

Below are some of the most **important design principles** associated with **DOE** that should be considered when planning your overall design plan.

Proper Sample Size & Power for a DOE

Remember, the most common analysis tool used with DOE is **ANOVA Analysis** which is a type of **hypothesis test** where we're **looking for differences in sample mean values** for different factors and their interactions.

So, it's important to refresh ourselves on the types of risks associated with a hypothesis test.

There are **two types of errors** in hypothesis testing.

The first error, **alpha risk**, is the risk that *the null hypothesis should not have been rejected and it was*, this is known as a **type I error**.

The second error, **beta risk**, is *when the null hypothesis should have been rejected and it wasn't*, this is known as a **type II error.**

The **probability** of a type II error is governed by the **beta risk (β)**, and it is analogous to the concept of consumers risk in the world of acceptance sampling.

Power = 1 – Beta Risk

Power is the probability of correctly rejecting the null hypothesis (H_0) when it is actually false.

Remember that we're only interested in the power of a hypothesis test **when the null hypothesis is in fact false**, which is when the various levels associated with our factor cause a statistically significant shift in the sample mean of our response.

When performing a **DOE,** we want to have **higher power**, which means **lower beta risk.**

How do we improve the power of our DOE? --- Increase your sample size!

Remember that within ANOVA we're analyzing sample means.

If you think back to the **inferential statistics** section, we learned that the **variance of the sample mean distribution** is a combination of the **population variance** and the **sample size**.

$$Variance\ of\ sample\ mean\ distribution: V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Increasing your sample size reduces the variability in your sample statistic distribution which improves your ability to discern between the null and alternative hypothesis when the null hypothesis is false – thus increasing **power**.

This is where **replication adds value to a DOE** in that it increases your sample size and **reduces beta risk and increases the power**.

Additional samples also reduce the **alpha risk** as well. In general, more samples help you make the right conclusion.

A Balanced Design

A **balanced design** is one where all of the treatments have the **same number of observations** or **replications**.

It is uncommon to have an unbalanced design and you should strive to ensure the same number of observations or replications per treatment.

Randomizing the Order of a Design

The **order of a design** refers to the **chronological sequence** in which you execute the various treatments within your design.

In general, the best designs are **ordered randomly**, in order to minimize the impact of uncontrollable factors.

Randomizing the order of a design ensures that the variation associated with the uncontrollable factors does not introduce any **bias** in the results.

Randomization can also apply to the way you allocate raw materials and other items to an experiment to ensure that any potential sources of variation are spread evenly across the design.

A design whose order of treatments is determined at random is considered a **completely randomized design**.

Blocking in DOE

Blocking is another method you can use to reduce the impact of uncontrollable factors on your experiment.

For example, let's say you knew that altitude was an uncontrollable factor in the cake baking experiment.

You could "block" for that factor by performing your experiment at the same altitude, thus eliminating the variation associated with that factor.

Or you could create two blocks, one block of experiments performed at sea level and one block of experiments performed in Denver (mile high).

Blocking lets you **minimize the variation** of an otherwise **uncontrollable factor** by carrying out your experiment at a single setting of that uncontrollable factor.

Blocking helps **reduce the experimental error** associated with our experiment, which increases the accuracy of the final ANOVA Analysis of the various factors and interactions.

There's a common saying in DOE – **Block** what you can, **randomize** what you can't.

A design where blocking has been used is called a **blocked design**.

If you combined a **random order with blocking,** you'll describe your design as **a completely randomized block design**.

An Efficient Design

An efficient design is one that **includes the minimal number of runs to accomplish the objective**.

In this way, you're **maximizing the value** associated with the time, effort and cost invested.

This is where being **clear about your objective** can save you time and effort. Why perform a full factorial when a fractional experiment will get you what you need?