# Chapter 24 – Design of Experiments

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

A **designed experiment** is one of 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.**

This chapter is laid out into 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, confounding** and **resolution**.

Section 4 is the **Full Factorial DOE** with an example**,** and section 5 is the **Fractional Factorial DOE**.

Section 6 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.

## 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 CQE Certification. These include Plackett-Burman Designs, Orthogonal Arrays and Response surface Designs (Central Composite Designs, Box-Behnken Designs).

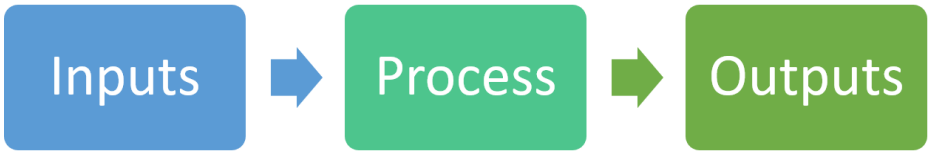
Lastly, this chapter is focused on **two-level designs**. Designs of three levels, or mixed-level designs are out of scope.

## 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

**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.

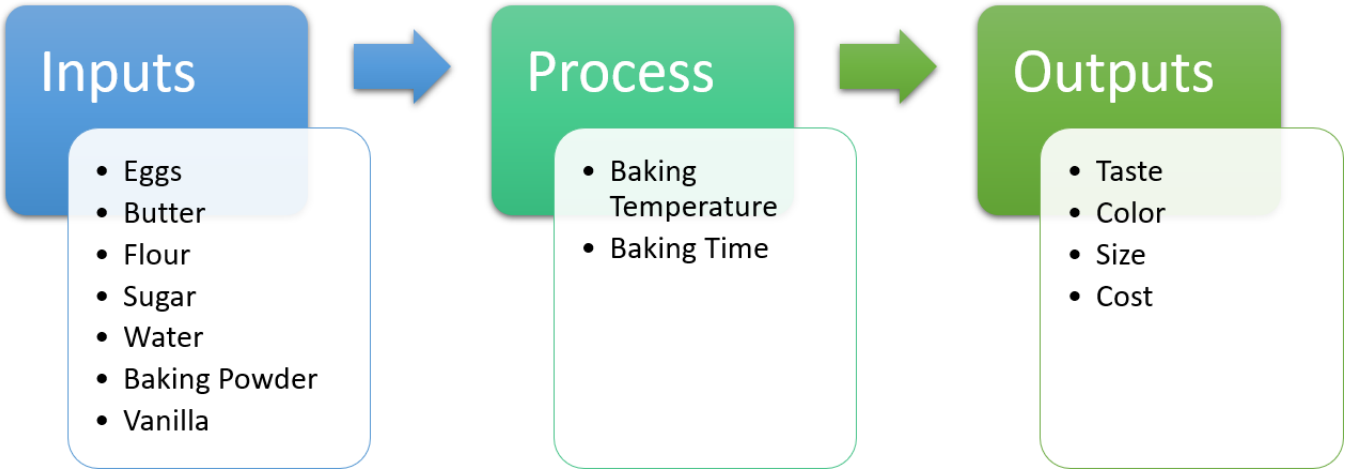
In the example of baking a cake there are factors that will influence the outputs but that are uncontrollable.

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)**.

## 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.

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.

| | 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 | - | - | - | - |

### *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 like 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.

If we go back to the analogy of the random, inherent process variation representing random error, then **systematic error** would be analogous with **special cause variation.**

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 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 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, is 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**.

## Repetition and Replication

Repetition and replication are fundamental principles within designed experiments, serving distinct yet interconnected roles in ensuring the validity of experimental results.

They contribute to reducing random variability and enhancing the precision of the study results.

**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.*

Replication provides the opportunity to assess the consistency of the treatment effect across different subjects or items.

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.

On the other hand, **Repetition** Involves applying the same treatment or condition to the same experimental unit multiple times.

It pertains to the number of measurements or observations taken within a single experimental unit under the same conditions.

## Nested Design

In a design experiment, "nested" refers to a situation where the levels of one factor are not interchangeable with or comparable to the levels of another factor.

Nesting in a design experiment involves a hierarchical structure where the levels of one factor exist within the levels of another factor.

It signifies a relationship where the subunits of one factor only exist within specific levels of another factor.

For example, in manufacturing, components produced within specific machines.

## Planning and Organizing a Designed Experiment

They're called **Designed Experiments not Haphazard Experiments** for a reason.

You have to be **intentional** about how you design and organize your experiments . . . if you want to get statistically valid results that is.

Oftentimes the **biggest mistakes in DOE are due to poor planning**, and thus importance of the planning phase cannot be understated.

Below is a 7-step process to plan and execute a DOE.

One last comment before we jump in: **DOE is a team sport.** In all of the steps below I'm going to assume that you're collaborating with the proper SME's in your organization.

### Step 1 – Determining the Experiment Objectives

The planning phase begins by determining the **objective of the experiment**.

The objective of your experiment **will dictate the most efficient design** to use. This will ensure that you've maximized the value for the effort you're going to put in.

There are **3 common "objectives"** or situations where a **DOE** is the right tool:

- Comparing Two Alternative Possibilities
- Screening/Characterizing a Process
- Modeling/Optimizing a Process

A **comparative DOE** is used when you want to make a comparison of factors at multiple levels. Usually this is a single factor, but can also include multiple factors.

An example of this would be a **comparative experiment** to study 2 different raw material vendors to determine the best one. Depending on the number of factors included in the study this can either be a **full or fractional factorial design**.

A **screening/characterization design** can be used to study your process as a whole to determine which factors are **critical** and which are not.

This type of design is typically a **fractional factorial design** due to the large number of **potentially critical factors** associated with your product or process. We will cover this type of design below.

A **modeling/optimization design** is meant study the **critical factors** associated with a product or process to **create a model of this process** and **determine the optimal levels of each factor**.

This type of design is usually a **full factorial design** as you've already been able to narrow your process down to 2-4 critical factors.

The optimization DOE can be used to hit a quality target, maximize yield, minimize variation, improve robustness, or optimize a process for various competing responses.

### Step 2 – Identifying Critical Responses and their Appropriate Measurement System

Step 2 and 3 in the DOE process is to **study your product/process** to understand all of the **potential inputs** and **outputs** that could be investigated.

A **process flow diagram** can be useful here in defining your process.

Once you know the purpose of your experiment, you must identify the **critical response variables – outputs**, that you want to measure.

You must also be confident that your **measurement system** for that response variable is **capable and stable**, as any variation within your measurement system increases your **systematic error** and will affect the final analysis.

You'd hate to get to the end of your experiment and find that your measurement system was introducing large amounts of variation on the response variable – which might render the experiment useless.

From a quality perspective, many of the process outputs should have been defined during the **product/process design phase**. These can include functional requirements, dimensional requirements, material requirements, etc.

From the **Continuous Improvement perspective**, critical responses can include process yield, process capacity, process cycle time, etc.

### Step 3 – Identifying Factors to Study

Once you're clear on the **critical responses (outputs)** that you want to study – it's time to **identify the factors (inputs)** that might influence your response variables.

The idea of a **screening design** is meant to do just this – define **critical and non-critical input factors** for your **responses**.

The other important consideration in this step is **determining appropriate levels for each factor** – the "Highs" and "Lows".

These highs and lows should be realistic and reasonable for your process, and should be allowed to vary to similar magnitudes.

We wouldn't try to bake a cake at 75° F or at 1,000° F.

Nor would we want to test **a very limited range** of baking temperatures like 370° F and 375° F. A range this narrow might indicate that temperature isn't a critical factor – which is simply incorrect.

Getting input during this stage from your **product/process SME's** is critical.

Many processes have hundreds of possible factors – it's important to focus your attention on only the critical few! This same comment goes for the response variables as well.

### Step 4 – Choosing the Right Design

This is the stage where you choose the proper design for your experiment. This should be based on the objective and the number of factors being studied.

If step 3 identified **7 input factors** to be studied - then you wouldn't perform a **full factorial design**.

Similarly, if step 3 identified only 2 factors, then you wouldn't want to perform a **fractional factorial design** – you'd want to go straight to the **full factorial**.

The other item to consider here are the **interactions**.

If you're simply attempting to identify the critical inputs in a **screening design**, then you're likely not worried about the interaction effects – thus you should likely pick a **fractional factorial design**.

As we will discuss below, when using a fractional factorial design, it's possible to experience **confounding results** between the **main effects** and the **interaction effects**, thus we typically omit the interactions from a screening experiment.

However, if you're attempting to **optimize or model your process**, you're going to want to understand the **interaction effects** between factors. In this case, the **full factorial design** is the one for you.

#### *Calculating the Number of Treatments for Two Level Designs*

This is a good time to explain how to calculate the number of treatments within a given design.

In a **full factorial experiment** the number of treatments is calculated as the **levels** raised to **factors** or $L^F$.

$$Full\ Factorial\ Design: Number\ of\ Treatments = Levels^{Factors} = L^F = 2^F$$

Since this chapter is focused on **2 level experiments**, I've shown that in the equation above.

In a **fractional factorial experiment,** the number of treatments is dependent on what fraction you want to use. The most **common fractions** are the **half (½) fraction** and the **quarter (¼) fraction**.

$$\boldsymbol{Half}\ Factorial\ Design: Number\ of\ Treatments = \frac{Levels^{Factors}}{2} = \frac{L^F}{2} = \frac{2^F}{2} = 2^{F-1}$$

$$\boldsymbol{Quarter}\ Factorial\ Design: Number\ of\ Treatments = \frac{Levels^{Factors}}{4} = \frac{L^F}{4} = \frac{2^F}{2^2} = 2^{F-2}$$

Below is a table of the **number of treatments** in the various **factorial design (Full, Half and Quarter)** that are required for different **number of factors**.

If your experiment had **7 inputs**, then to perform a **full factorial** experiment means you'd have to perform **128 treatments**. Yikes. If you chose a **quarter factorial** design, you'd only have to run **32 treatments**, much better.

| # of Factors | Full Factorial Exp. | Half Factorial Exp. | Quarter Factorial Exp. |
|---|---|---|---|
| **2** | 4 | 2 | 1 |
| **3** | 8 | 4 | 2 |
| **4** | 16 | 8 | 4 |
| **5** | 32 | 16 | 8 |
| **6** | 64 | 32 | 16 |
| **7** | 128 | 64 | 32 |
| **8** | 256 | 128 | 64 |
| **9** | 512 | 256 | 128 |
| **10** | 1024 | 512 | 256 |

### Step 5 – Run the Experiment

Ok, so once you've identified your inputs, outputs and the proper design, it's time to **execute** the experiment.

This is always considered the **expensive part of the experiment** – because it's where the majority of your time and money will be consumed – so any **mistakes** in this phase will be **costly**.

This is why it's important to have invested the right amount of time, thought and energy into the **planning phase**, to avoid mistakes here.

One key thing to keep in mind is to **collect data** in a way that **minimizes the chance for error**. A **check sheet** can be a powerful tool here.

If **human bias** is a concern – attempt to design a "blind" study where the levels associated with each treatment are not known to the humans doing the data collection.

Lastly, make sure to **run the experiment in the proper order** – perhaps in a **completely randomized fashion**. More on this later!

### Step 6 – Analyze the Results

There are a handful of ways to analyze the results of a DOE. With a simple one factor experiment at two levels you can simply perform a **t-test (hypothesis test)**.

With a more complex experiment, **ANOVA** is the preferred method, which we will review below.

In today's world software packages exist to both create and analyze DOE's, so much of the heavy lifting is done for you. However, you still must understand the common concepts and tools within DOE to ensure you're maximizing your results.

### Step 7 – Make Decisions, Iterate and Plan your Next Experiment

Once you've analyzed your results it's time to act. Use this new-found knowledge to improve your process.

Or, if the goal of your first experiment was to screen out the critical factors (screening design), you can use this new knowledge to start planning another DOE!

In fact, it's worth noting that often the best DOE approach is the **iterative approach**.

A lot of the best experiments starts with a screening design (**fractional factorial**) to determine the critical factors, then an optimization design (**full factorial**) to optimize your process/product.

It is usually better to perform 2 or 3 smaller experiments that all build off of each other - than 1 large experiment.

When experiments are planned in an iterative way, the knowledge gained from the first experiment can be used to change and improve subsequent experiments, allowing the knowledge to compound over time.

This can include **replicating** prior results.

This is like the PDCA cycle – which is an in iterative cycle of learning, experimentation and implementation.

## 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 the treatments have the **same number of observations** or **replications**.

### 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.

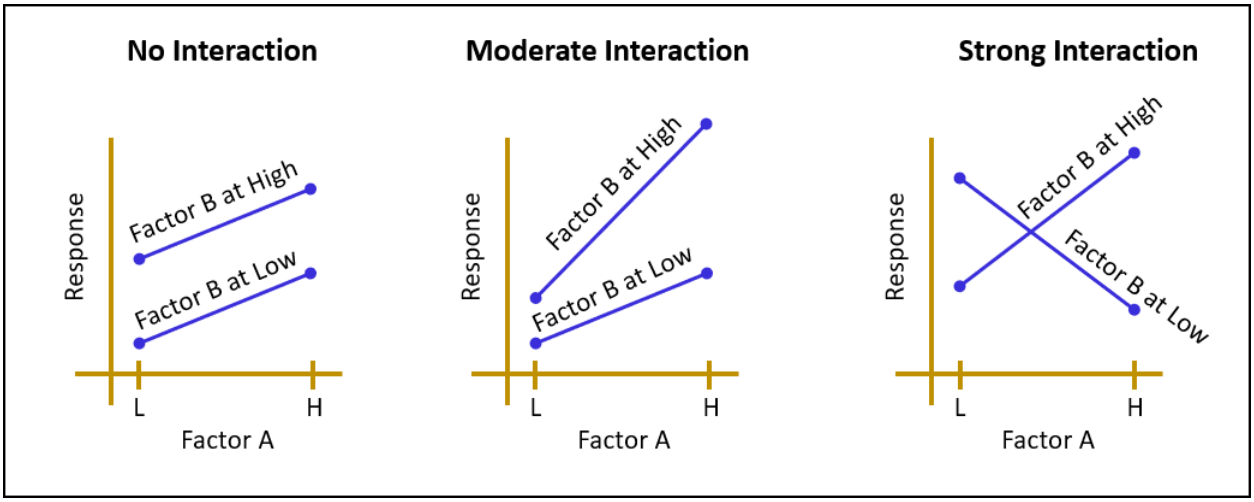
This is where being **clear about your objective** can save you time and effort. Why perform a full factorial when a fractional experiment will get you what you need?

## Interactions Between Factors

When an experiment has multiple factors, often **two input factors can "interact"** in a way where they **simultaneously influence the response variable**.

Two factors are said to have an interaction when the **response variable changes** when **both factors** are varied **simultaneously**.

Below are some examples of what **interactions** look like between two factors (A & B).




The far-right image is the best example.

When Factor B is at the high setting, the response variable increases when Factor A moves from low to high.

When Factor B is at the Low setting though, the response variable decreases when Factor A moves from low to high.

Do you see how Factor B interacts with Factor A to affect the response?

**Interactions** can be fully analyzed in a **full factorial experiment** where **all possible combinations of levels and factors** are studied.

Remember if you do want to study the **interaction effects**, then **replication** (more samples), might be needed to ensure there will **be enough degrees of freedom** to analyze the effects of all possible **interactions**.

Oftentimes in a fractional factorial experiment the **interactions will be confounding** with the **main effects** of a given factor and thus cannot be examined. More on this below.

The last comment worth making about **interactions** is that this is **where DOE's often add the most value**. Interactions cannot be observed when performing an OFAT (One Factor at a Time) experiment, and can only be observed during a DOE.

## Confounding Results

**Factors can be described as Confounding when the effect on the response variable cannot be separated into causal relationships for each factor.**

Two factors are confounding when their effects are indistinguishably combined to affect the response variable.

**Confounding** can often be the result of poor planning in the design phase, when factors are varied in similar ways such that the change in the response variable cannot be attributed to a single factor.

**Confounding often occurs** when dealing with the **interactions** of a **fractional factorial design** where only a limited number of runs are executed. This is why the interactions are often ignored during a fractional factorial study.

| | | Factors | | | | Interactions | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Factor A | Factor B | Factor C | Factor D | AB | AC | AD | BC | BD | CD | Results |
| Treatments | 1 | + | + | + | + | + | + | + | + | + | + | Good |
| | 3 | + | - | - | + | - | - | + | + | - | - | Bad |
| | 5 | - | + | - | + | - | + | - | - | + | - | Bad |
| | 7 | - | - | + | + | + | - | - | - | - | + | Good |
| | 9 | + | + | + | - | + | + | - | + | - | - | Good |
| | 11 | + | - | - | - | - | - | - | + | + | + | Bad |
| | 13 | - | + | - | - | - | + | + | - | - | + | Bad |
| | 15 | - | - | + | - | + | - | + | - | + | - | Good |

There are 3 different instances of confounding results within this experiment. Take the color purple, notice how Factor C, and the Interaction between A and B have the same experimental design (+ - - + - - +).

And also notice how the results (Good and Bad) are always good when Factor C is high (+) and always bad when Factor C is low (-), but that comment is also true for the interaction between A and B?

This design was planned poorly (for illustration purposes 😊), and now **our results are confounding**.

That is, **we can't distinguish or separate the effects on taste into a causal relationship with each factor.**

To reduce the amount of confounding results, we can add additional treatment groups to our experiment and improve the resolution of our design.

## The Resolution of your Designed Experiment

Resolution in the context of designed experiments, particularly in fractional factorial designs, refers to the degree to which interactions between factors are confounded or aliased with one another.

Resolution is denoted by a numerical value (e.g., Resolution III, Resolution IV) and is typically designated by Roman numerals. Different types of resolution include:

**Resolution III designs**: These designs confound main effects with two-factor interactions, meaning they don't allow for the estimation of two-factor interactions without ambiguity. They're suitable for screening experiments where identifying which main effects are significant is the primary goal.

**Resolution IV designs:** These designs confound main effects with three-factor interactions, allowing for the estimation of main effects without ambiguity. However, two-factor interactions remain confounded. They're used for more detailed investigations following a screening experiment and can estimate a wider range of interactions.

**Higher Resolutions:** There are higher resolution designs (such as V and above), which further reduce the confounding of interactions.

Resolution V designs, for instance, confound up to four-factor interactions but provide greater precision in estimating main effects and lower-order interactions.

The choice of resolution in a designed experiment depends on the goals, resources, and the level of detail required to address the research questions.

**Higher resolutions offer more detailed information but often require more runs, making them more resource-intensive.**

Designers balance the trade-off between precision in estimating effects and the number of experimental runs needed.

**Lower resolution designs, while more efficient in terms of runs, might sacrifice precision in estimating interactions.**

## One Factor Experiments

A **completely randomized design** in experimental design (DOE) involves randomly assigning treatments to experimental units without any pre-defined structure.

Essentially, you're using the concept or **randomization** when determining the sequence of your experiment.

This method ensures randomness and minimizes bias, allowing for the analysis of treatment effects while assuming the independence of observations.

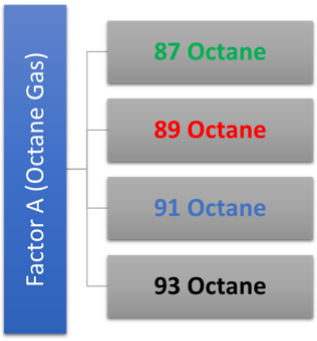
| | Critical Factors | Replicates of the Response (Horsepower) | | | |
|---|---|---|---|---|---|
| **Runs** | **Octane** | **1** | **2** | **3** | **4** |
| Treatment 1 | 87 | 8 | 11 | 10 | 3 |
| Treatment 2 | 89 | 16 | 1 | 14 | 4 |
| Treatment 3 | 91 | 7 | 9 | 2 | 12 |
| Treatment 4 | 93 | 6 | 5 | 15 | 13 |

It's a straightforward design, ideal for studying a single factor offering simplicity and statistical validity in comparing treatment effects.

A **completely randomized, blocked design** in experimental design (DOE) is essentially any design that involves the usages of randomization, and blocking to reduce or eliminate error in an experiment to increase the validity of the results of the experiment.

This style of design normally includes a single factor that is being studied, along with multiple blocking factors. These blocking factors can be varies from high to low, to measure the impact of those blocking factors.

| | Critical Factors | Blocking Factor | Replicates of the Response (Horsepower) | | | |
|---|---|---|---|---|---|---|
| **Runs** | **Octane** | **Mileage** | **1** | **2** | **3** | **4** |
| Treatment 1 | 87 | 0 Miles | | | | |
| Treatment 2 | 89 | 0 Miles | | | | |
| Treatment 3 | 91 | 0 Miles | | | | |
| Treatment 4 | 93 | 0 Miles | | | | |
| Treatment 5 | 87 | 100K Miles | | | | |
| Treatment 6 | 89 | 100K Miles | | | | |
| Treatment 7 | 91 | 100K Miles | | | | |
| Treatment 8 | 93 | 100K Miles | | | | |




**We want to block out Mileages as Factor**

## Latin Squares

A Latin Square Design is a unique DOE that you can use when you have **a single factor**, that you want to vary at multiple levels, and you also have 2 blocking factors.

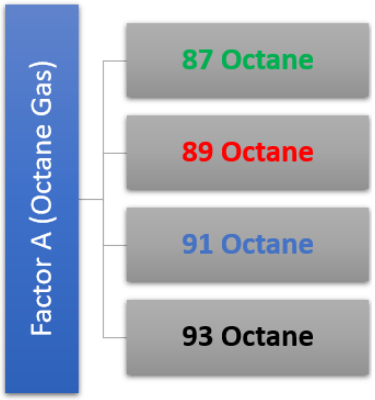
The key assumption in Latin squares is that there is **no interaction between the primary factor that you want to study**, and the two blocking factors that are also included in the experiment.

For example, if we wished to study octane gas (factor) across 4 levels (87, 89, 91 and 93 octane).

We could create a **4x4 Latin square**, where we also include **two blocking factors** (brand of vehicle and mileage of vehicle).

Another key factor about the Latin square, is that the number of levels associated with your primary factor (4x), must be mirrored by the number of levels in your blocking factors.

| | | **Blocking Factor A (Brand of Vehicle)** | | | |
|---|---|---|---|---|---|
| | | Level 1 | Level 2 | Level 3 | Level 4 |
| **Blocking Factor B (Mileage)** | Level 1 (0 Miles) | 87 | 89 | 91 | 93 |
| | Level 2 (25K Miles) | 89 | 91 | 93 | 87 |
| | Level 3 (50K Miles) | 91 | 93 | 87 | 89 |
| | Level 4 (100K Miles) | 93 | 87 | 89 | 91 |




And if you wanted to only study 3 levels of the single factor (octane gas), you would have to switch to a 3x3 latin square and reduce the number of levels of your blocking factor to 3 levels.

| | | **Blocking Factor A (Brand of Vehicle)** | | |
|---|---|---|---|---|
| | | Level 1 | Level 2 | Level 3 |
| **Blocking Factor B (Mileage)** | Level 1 (0 Miles) | 87 | 89 | 91 |
| | Level 2 (25K Miles) | 89 | 91 | 87 |
| | Level 3 (50K Miles) | 91 | 87 | 89 |

## Full-Factorial Design of Experiments

Alright, on to one of the most fundamental and common designs within DOE – the **full factorial design**.

I'll start this with a brief explanation of the details of a full factorial design, then we will jump into an **example** of a 3-factor design at 2 levels.

A **full factorial experiment** is one in which **every combination of factors and levels** is included within the experiment.

The full factorial experiment is a good design if you're attempting to **model or optimize your process**, and when you're looking to analyze both the **main effects of each factor** along with any possible **interactions between factors**.

Remember, in this design, **all combinations of factors at the various levels** are included in the experiment. So, the more factors you pick, the more experiments you have to run.

| # of Factors | Full Factorial Exp. | Half Factorial Exp. | Quarter Factorial Exp. |
|---|---|---|---|
| **2** | 4 | 2 | 1 |
| **3** | 8 | 4 | 2 |
| **4** | 16 | 8 | 4 |
| **5** | 32 | 16 | 8 |
| **6** | 64 | 32 | 16 |
| **7** | 128 | 64 | 32 |
| **8** | 256 | 128 | 64 |
| **9** | 512 | 256 | 128 |
| **10** | 1024 | 512 | 256 |

If you have 6 factors, you'll end up running **64 different treatment** groups, if you have 9 factors, you're running **512 treatments**.

### Example of a Full Factorial Experiment

Let's use an example to demonstrate how we would **construct a full factorial design**, and discuss how you can **graphically and computationally calculate the main effects and interaction effects**.

Here's the **experiment**:

You manage a process where you know that **3 factors** (temperature, pressure and time) influence the **response variable** (yield) associated with the process, and you want to **maximize the yield**.

Recall that we're planning a **2-level** experiment, where a **high and low level** will be defined for each **factor** which are commonly shown as **+ (high) and – (low)**.

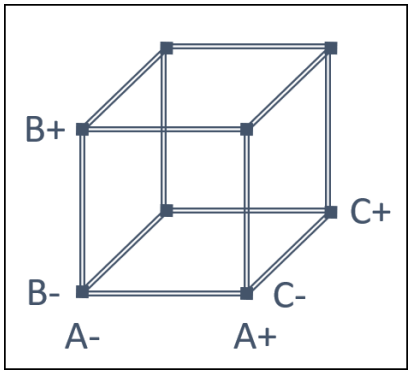
You might also see these high and low levels defined as **+1 or -1**.

Below is a design matrix for this 3-factor, 2 level design, where you can see how **the design requires 8 treatments to capture all of the unique combination of the levels of each factor.**

Treatment #1, is the experiment where all factors will be set at "**high**", and treatment 8 is the experiment where all factors will be set at "**low**".

| | | Factors | | |
|---|---|---|---|---|
| | | Temperature | Pressure | Time |
| Treatments | 1 | + | + | + |
| | 2 | + | + | - |
| | 3 | + | - | + |
| | 4 | + | - | - |
| | 5 | - | + | + |
| | 6 | - | + | - |
| | 7 | - | - | + |
| | 8 | - | - | - |

One way you visually represent these 8 treatments is with a cube showing the different high and low conditions for the 3 factors.




The 8 corners of this cube represent the 8 treatments, each of which is a unique combination of the high and low levels for each of the 3 factors (A, B & C).

In discussion with our subject matter experts, we want to try to following **high** and **low levels** for each **factor**:

| Factor | High | Low |
|---|---|---|
| **Temperature** | 270° F | 240° F |
| **Pressure** | 400psi | 300psi |
| **Time** | 45 sec | 30 sec |

Let's execute our experiment and see what kind of results we get, then we can review the **computational and graphical methods for analyzing these results**.

| | | Factors | | | Response |
|---|---|---|---|---|---|
| | | Temperature | Pressure | Time | % Yield |
| Treatments | 1 | + | + | + | 83 |
| | 2 | + | + | - | 75 |
| | 3 | + | - | + | 87 |
| | 4 | + | - | - | 95 |
| | 5 | - | + | + | 95 |
| | 6 | - | + | - | 52 |
| | 7 | - | - | + | 89 |
| | 8 | - | - | - | 78 |

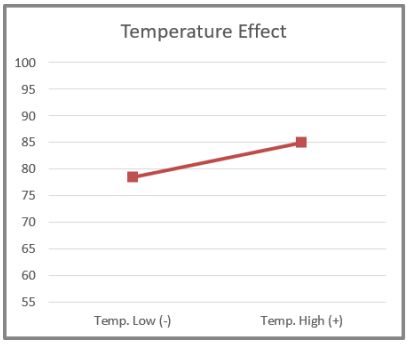
Let's start by **calculating and graphing** the **main effects**, then move on and **calculate the interactions between factors**.

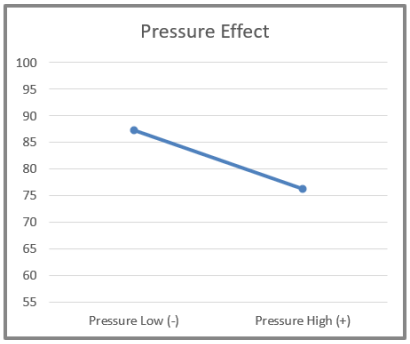
## Calculating and Graphing the Main Effects

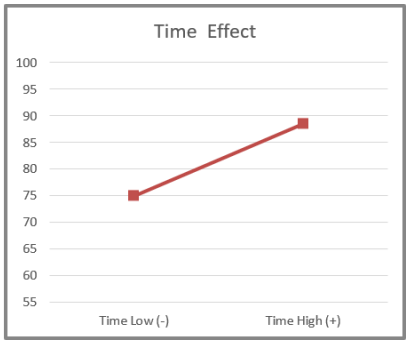
Like we discussed above, **ANOVA Analysis tool** is a great to determine if the levels of each factor have a statistically significant effect on the response variable.

You can also use a graphical method and computational method to **determine the effect of each factor on the response variable**.

Let's start with the graphical method by looking at the **main effects plots**.




The **main effects plots** will help you **visualize the effect of each factor** at each **level** (high and low).

This graph shows the **average response value** (Yield in our case) at the **two levels for each factor,** and how the **response changes** as you move from low to high.

For temperature, treatments 1 – 4 would be averaged for the high value, and treatment 5-8 would be averaged for the low value.

For pressure, treatments 1, 2, 5 & 6 would be averaged for the high value, and treatment 3,4, 7 & 8 would be averaged for the low value.

Make sense?

There's also a **computational method to calculate the estimated effect of each factor at the two levels.**

This **computational method** is simply the difference in the average value at the high level minus the average value at the low level, which is exactly what we graphed above in the **main effect plots**:

$$\boldsymbol{Estimated\ Effect = Average\ at\ High - Average\ at\ Low}$$

Let's use our experimental data to calculate the estimated effect on the yield for each factor.

$$\boldsymbol{Temperature}\ Estimated\ \boldsymbol{Main\ Effect} = \frac{83 + 75 + 87 + 95}{4} - \frac{85 + 52 + 89 + 78}{4} = 6.5$$

$$\boldsymbol{Pressure}\ Estimated\ \boldsymbol{Main\ Effect} = \frac{83 + 75 + 95 + 52}{4} - \frac{87 + 95 + 89 + 79}{4} = -11$$

$$\boldsymbol{Time}\ Estimated\ \boldsymbol{Main\ Effect} = \frac{83 + 87 + 95 + 89}{4} - \frac{75 + 95 + 52 + 78}{4} = 13.5$$

Based on this data we can see that **Time has the biggest effect on the yield**, and that a longer time value could potentially increase the yield by 13.5 points.

**Pressure has a strong negative effect on yield** where the low setting yielded 11% better results than the high setting.

**Temperature had the smallest effect on yield**, with a high temperature resulting in 6.5% more yield than the low temperature.

Before we make any conclusions however, we should check out the interactions between the factors.

## Calculating and Graphing the Interaction Effects

Let's go back to our original design matrix to understand how we **calculate a "high" or "low" level for each interaction effect**.

| | | Factors | | | Interactions | | | Response |
|---|---|---|---|---|---|---|---|---|
| | | Temperature | Pressure | Time | Temperature x Pressure | Temperature x Time | Pressure x Time | % Yield |
| Treatments | 1 | + | + | + | + | + | + | 83 |
| | 2 | + | + | - | + | - | - | 75 |
| | 3 | + | - | + | - | + | - | 87 |
| | 4 | + | - | - | - | - | + | 95 |
| | 5 | - | + | + | - | - | + | 95 |
| | 6 | - | + | - | - | + | - | 52 |
| | 7 | - | - | + | + | - | - | 89 |
| | 8 | - | - | - | + | + | + | 78 |

The high and low value for each factor is combined to determine if the interaction level is a high or low.

This is easiest when you imagine the **+ as a +1**, and the **– as a -1**. Then you're simply multiplying the two values together.

So, when temperature and pressure are both at high (+), the interaction is also at a high (+1 * +1 = +1). When Temperature is at a high (+), but time is at a low (-), the resulting interaction is at a low (1 * -1 = -1).

Once we have the high and low values for each interaction, we can calculate the **estimated effect** for that **interaction** in the same way we did above for the main effect.

$$\boldsymbol{Estimated\ Effect = Average\ at\ High - Average\ at\ Low}$$

Let's calculate the interaction effect between our 3 factors:

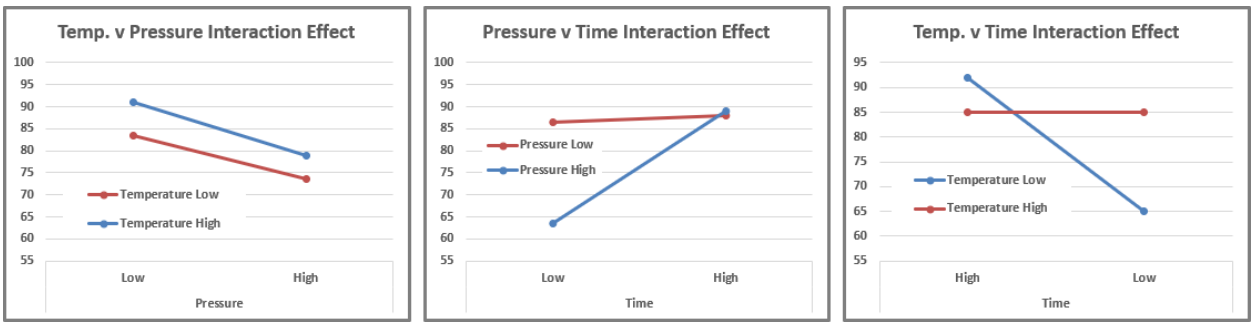
$$\boldsymbol{Temp.\ \&\ Press.\ Interaction\ Effect} = \frac{83+75+89+78}{4} - \frac{87+95+95+52}{4} = -1$$

$$\boldsymbol{Pressure\ \&\ Time\ Interaction\ Effect} = \frac{83+95+95+78}{4} - \frac{75+87+52+89}{4} = 12$$

$$\boldsymbol{Temp.\ \&\ Time\ Interaction\ Effect} = \frac{83+87+52+78}{4} - \frac{75+95+95+89}{4} = -13.5$$

So, you'll notice that there are some **strong interaction** effects here, especially related to **time**.

Let's see what these interactions look like in an **interaction effects plot:**




Within this example, we've only **analyzed the first order interactions** – which are the single combinations of interactions between each main factor (Pressure and Time, etc).

We are going to ignore the **second order interactions**, which in this situation would be the combination of all 3 factors combined (Temperature X Pressure X Time).

In this scenario, with 3 factors, there is only **a single second order interaction**. However, when you have more than 3 factors, the number of second order interactions grows, along with **third and fourth order interaction effects**.

At the end of this experiment, it looks like the **most favorable combination of factors is treatment 4**, which maximizes the yield, and reduces the overall time associated with the process, which also impacts the capacity & throughput of the process.

To confirm this result, **replication can be used to ensure that the results are consistent over time**, **and ANOVA analysis** can be used to ensure the results are statistically significant.

As you can see, a full factorial DOE can be very powerful, however if the number of factors involved in the DOE grows the effort required to conduct the experiment grows exponentially with the number of required treatments.

For example, if you had a process with 9 possibly critical variables, a full factorial DOE would require 512 treatments. yikes.

To combat this issue, statisticians have developed the factorial DOE to allow only us to execute only a portion of the treatments and still make reasonable conclusions.

## Fractional Factorial Design of Experiments

As the name implies, a fractional factorial experiment is an experiment where only a fraction of the possible treatments are conducted.

Let's review the info from above to see the benefit of a fractional experiment.

In a **full factorial experiment** the number of treatments is calculated as the **levels** raised to **factors** or $L^F$.

$$Full\ Factorial\ Design: Number\ of\ Treatments = Levels^{Factors} = L^F = 2^F$$

In a **fractional factorial experiment,** the number of treatments is dependent on what fraction you want to use.

The most **common fractions** are the **half (½) fraction** and the **quarter (¼) fraction**.

$$\boldsymbol{Half}\ Factorial\ Design: Number\ of\ Treatments = \frac{Levels^{Factors}}{2} = \frac{L^F}{2} = \frac{2^F}{2} = 2^{F-1}$$

$$\boldsymbol{Quarter}\ Factorial\ Design: Number\ of\ Treatments = \frac{Levels^{Factors}}{4} = \frac{L^F}{4} = \frac{2^F}{2^2} = 2^{F-2}$$

Below is a table of the **number of treatments** in the various **factorial design (Full, Half and Quarter)** that are required for different **number of factors**.

If your experiment had **7 inputs**, then to perform a **full factorial** experiment means you'd have to perform **128 treatments**. Yikes. If you chose a **quarter factorial** design, you'd only have to run **32 treatments**, much better.

| **# of Factors** | Full Factorial Exp. | Half Factorial Exp. | Quarter Factorial Exp. |
|---|---|---|---|
| **2** | 4 | 2 | 1 |
| **3** | 8 | 4 | 2 |
| **4** | 16 | 8 | 4 |
| **5** | 32 | 16 | 8 |
| **6** | 64 | 32 | 16 |
| **7** | 128 | 64 | 32 |
| **8** | 256 | 128 | 64 |
| **9** | 512 | 256 | 128 |
| **10** | 1024 | 512 | 256 |

Let's take an example where we had 4 factors, however we could only afford to run 10 experimental runs (treatments).

We would have to limit ourselves to a half factorial experiment, where we'd end up running 8 treatments (experimental runs).

This means we'd be running **a half factor experiment**, so let's look at what a possible **design matrix** would look like in this example.

| Treatments | Factor A | Factor B | Factor C | Factor D |
|---|---|---|---|---|
| 1 | + | + | + | + |
| 2 | + | + | - | + |
| 3 | + | - | - | + |
| 4 | + | - | + | + |
| 5 | - | + | - | + |
| 6 | - | + | + | + |
| 7 | - | - | + | + |
| 8 | - | - | - | + |
| 9 | + | + | + | - |
| 10 | + | + | - | - |
| 11 | + | - | - | - |
| 12 | + | - | + | - |
| 13 | - | + | - | - |
| 14 | - | + | + | - |
| 15 | - | - | + | - |
| 16 | - | - | - | - |

Full Factorial Design

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

Fractional Factorial Design

What I've shown below is the full factorial experiment on the left, and I've highlighted in red the treatments that we could exclude from the fractional factorial design.

Remember, when executing a fractional factorial study, **you must be aware of the possibility of confounding** between main effects and the interaction effects.

Below are the first order interactions between these 4 factors, where I've highlighted the **confounding situations between the main effects and the interactions**.

For example, the **interaction between factor A and B** (AB) is **confounding with the main effect of factor C**, and the main effect of factor A is confounded with the interaction between B & C.

| | | Factors | | | | Interactions | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Factor A | Factor B | Factor C | Factor D | AB | AC | AD | BC | BD | CD |
| Treatments | 1 | + | + | + | + | + | + | + | + | + | + |
| | 3 | + | - | - | + | - | - | + | + | - | - |
| | 5 | - | + | - | + | - | + | - | - | + | - |
| | 7 | - | - | + | + | + | - | - | - | - | + |
| | 9 | + | + | + | - | + | + | - | + | - | - |
| | 11 | + | - | - | - | - | - | - | + | + | + |
| | 13 | - | + | - | - | - | + | + | - | - | + |
| | 15 | - | - | + | - | + | - | + | - | + | - |

Recall that the fractional design is a good choice when you're attempting to **screen out the critical factors** from non-critical factors, and thus the **interaction effects are less important**.

You'll often hear these fractional designs called "**main effects designs**" because they only seek to assess the main effects of each factor and not the interactions.

The **analysis of the main effects** within a **fractional factorial design** is the same as a full factorial design, it can be done both **graphically or computationally**.