DOE Terminology

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

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

The **outputs** are also commonly called **Responses** or **Dependent Variables**. **Outputs** (**response variables**) represent the **outcome** of a process or experiment.

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

A l**evel** refers to **specific settings** of a **factor (Input)**, and a **treatment** is a **unique combination of factors and levels** within an experiment**.**

During an experiment you might observe that your **response variable** experiences variation that **cannot be explained by the experiment** and the **factors** that you've decided to vary. **This unexplainable variation in your response (output) is called experimental error**.

There are **two types** of **experimental error** that you must be aware of - **random error** and **systematic error**.

Random error is the variation caused by **uncontrollable factors (noise)** or random variation in the response variable.

There is another type of error that is **systematic in nature**, and is not related to the natural variation in your response variable. The classic example of systematic nature is **measurement bias** or **measurement error**.

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

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**.

Robustness is the degree to which a product or process is unaffected by the variation of a particular factor – usually an uncontrollable factor**.**

Planning & Organizing Experiments

The **planning phase of DOE** begins by determining the **objective of the experiment**, and there are **3 common "objectives"** or situations where a **DOE** is the right tool.

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.

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

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**.

Design Principles

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

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

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

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

Blocking lets you **minimize the variation** of an otherwise **uncontrollable factor** by carrying out your experiment at a single setting of that uncontrollable factor. 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 is one that **includes the minimal number of runs to accomplish the objective**.

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**. **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**.

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.

Full-Factorial Design of Experiments

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

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**.

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

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

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**.

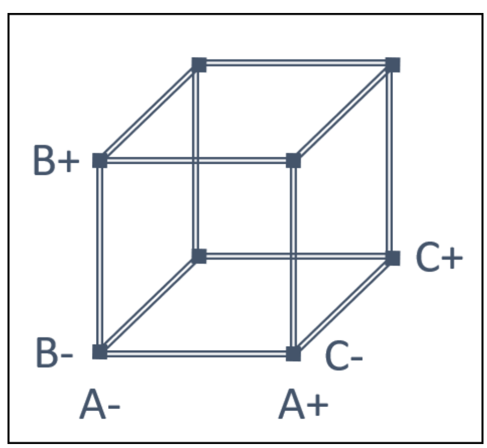
# of Factors	Full Factorial Exp.	Half Factorial Exp.	Quarter Factorial Exp.
2	4	2	1
3	8	4	2
4	16	8	4
5	32	16	8
6	64	32	16
7	128	64	32
8	256	128	64
9	512	256	128
10	1024	512	256

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.

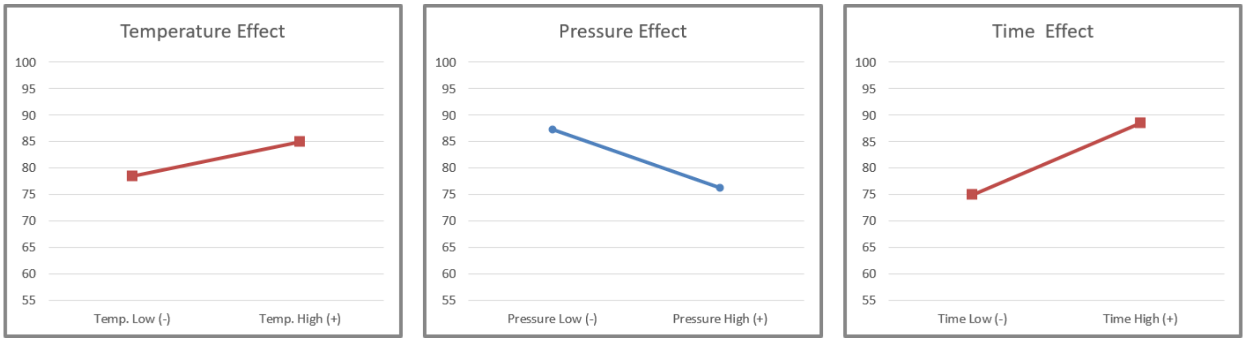
Below is a design matrix for this 3-factor, 2 level design, where you can see how **the design requires 8 treatments to capture all of the unique combination of the levels of each factor.**

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	-	-	-



Let's start with the **graphical method** for **analyzing DOE Results** 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.

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**:

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

Interactions

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.

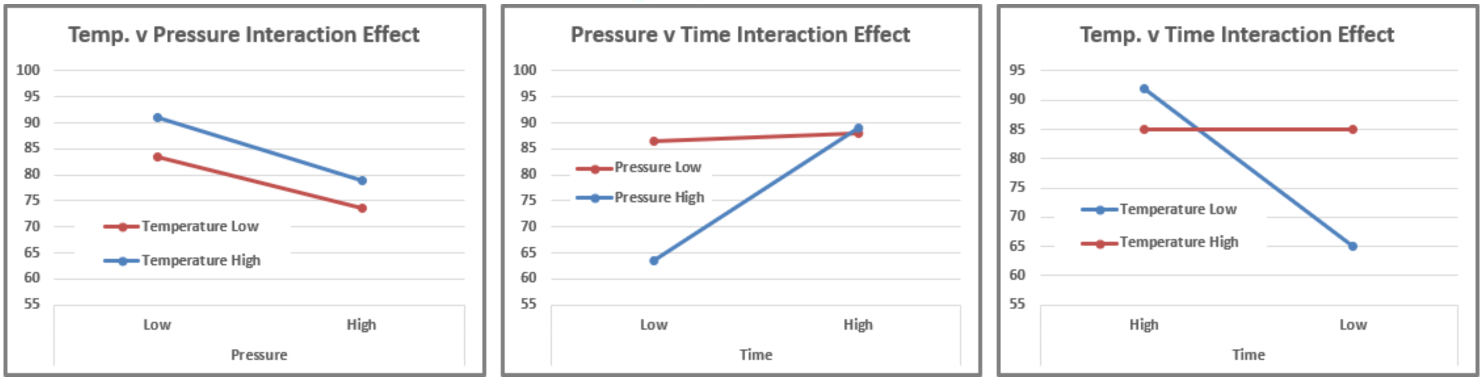
This is easiest when you imagine the **+ as a +1**, and the **– as a -1**. Then you're simply multiplying the two values together.

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



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.

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**.

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

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

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

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**.