Sub Module 1.6

6. Design of experiments

Goal of experiments:

- *Experiments help us in understanding the **behavior** of a (mechanical) system*
- *Data collected by systematic variation of influencing **factors** helps us to quantitatively describe the underlying phenomenon or phenomena*

The goal of any experimental activity is to get the **maximum** information about a system with the **minimum** number of well designed experiments. An experimental program recognizes the major **"factors"** that affect the outcome of the experiment. The factors may be identified by looking at all the quantities that may affect the outcome of the experiment. The most important among these may be identified using a few **exploratory** experiments or from **past experience** or based on some underlying **theory** or **hypothesis**. The next thing one has to do is to choose the number of **levels** for each of the factors. The data will be gathered for these values of the factors by performing the experiments by maintaining the levels at these values.

Suppose we know that the phenomena being studied is affected by the pressure maintained within the apparatus during the experiment. We may identify the smallest and the largest possible values for the pressure based on experience, capability of the apparatus to withstand the pressure and so on. Even though the pressure may be varied **"continuously"** between these limits, it is seldom necessary to do so. One may choose a few values within the identified **range** of the pressure. These will then be referred to as the levels.

Experiments repeated with a particular set of levels for all the factors constitute replicate experiments. Statistical validation and repeatability concerns are answered by such replicate data.

In summary an experimental program should address the following issues:

- *Is it a single quantity that is being estimated or is it a trend involving more than one quantity that is being investigated?*
- *Is the trend linear or non-linear?*
- *How different are the influence coefficients?*
- *What does dimensional analysis indicate?*
- *Can we identify dimensionless groups that influence the quantity or quantities being measured*
- *How many experiments do we need to perform?*
- *Do the factors have independent effect on the outcome of the experiment?*
- *Do the factors interact to produce a net effect on the behavior of the system?*

Full factorial design:

A full factorial design of experiments consists of the following:

- *Vary **one factor** at a time*
- *Perform experiments for **all levels** of all factors*
- *Hence perform a large number of experiments that are needed!*
- *All **interactions** are captured (as will be shown later)*

Consider a simple design for the following case:

Let the number of factors = k

Let the number of levels for the i^{th} factor = n_i

The total number of experiments (n) that need to be performed is $n = \prod_{i=1}^{k} n_i$.

If k = 5 and number of levels is 3 for each of the factors the total number of experiments to be performed in a full factorial design is $3^5 = 243$.

2^k factorial design:

Consider a simple example of a 2^k factorial design. Each of the k factors is assigned only two levels. The levels are usually **High = 1** and **Low = -1**. Such a scheme is useful as a preliminary experimental program before a more ambitious study is undertaken. The outcome of the 2^k factorial experiment will help identify the relative importance of factors and also will offer some knowledge about the interaction effects. Let us take a simple case where the number of factors is 2. Let these factors be x_A and x_B. The number of experiments that may be performed is 4 corresponding to the following combinations:

Experiment No.	x_A	x_B
1	+1	+1
2	-1	+1
3	+1	-1
4	-1	-1

Let us represent the outcome of each experiment to be a quantity y. Thus y_1 will represent the outcome of experiment number 1 with both factors having their "**High"** values, y_2 will represent the outcome of the experiment number 2 with the factor A having the **"Low"** value and the factor B having the **"High"** value and so on. The outcome of the experiments may be represented as the following matrix:

$x_A \downarrow \quad x_B \rightarrow$	+1	-1
+1	y_1	y_3
-1	y_2	y_4

A simple **regression model** that may be used can have up to **four** parameters. Thus we may represent the regression equation as

$$y = p_0 + p_A x_A + p_B x_B + p_{AB} x_A x_B \qquad \textbf{(48)}$$

The p's are the parameters that are determined by using the **"outcome"** matrix by the simultaneous solution of the following four equations:

$$\begin{aligned} p_0 + p_A + p_B + p_{AB} &= y_1 \\ p_0 - p_A + p_B - p_{AB} &= y_2 \\ p_0 - p_A - p_B - p_{AB} &= y_3 \\ p_0 - p_A - p_B + p_{AB} &= y_4 \end{aligned} \qquad \textbf{(49)}$$

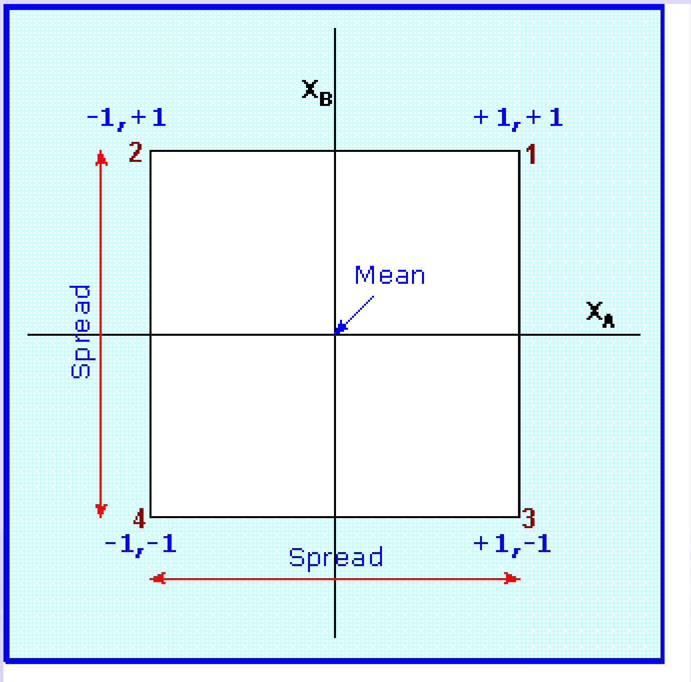

Figure 14 Interpretation of 2^2 factorial experiment

It is easily seen that the parameter p_0 is simply the mean value of y that is obtained by putting $x_A = x_B = 0$ corresponding to the mean values for the factors. Equation (49) expresses the fact that the outcome may be interpreted as shown in Figure 14.

It is thus seen that the values of y- p_0 at the corners of the square indicate the **deviations** from the mean value and hence the mean of the square of these deviations (we may divide the sum of the squares with the number of degrees of freedom = 3) is the **variance** of the **sample data collected** in the experiment. The influence of the factors may then be gauged by the contribution of each term to the variance. These ideas will be brought out by example 13.

Example 13

A certain process of finishing a surface involves a machine. The machine has two speed levels x_A and the depth of cut x_B may also take on two values. The two values are assigned +1 and -1 as explained in the case of 2^2 factorial experiment. The outcome of the process is the surface finish y that may have a value between 1 (the worst) to 10 (the best). A 2^2factorial experiment was performed and the following matrix gives the results:

$x_A \downarrow$ $x_B \rightarrow$	+1	-1
+1	3.5	1.5
-1	8.2	2

Determine the regression parameters and comment on the results. The regression model given in equation (48) is made use of. The four simultaneous equations for the regression parameters are given by

$$
\begin{aligned}
p_0 + p_A + p_B + p_{AB} &= 3.5.....(i) \\
p_0 - p_A + p_B - p_{AB} &= 1.5.....(ii) \\
p_0 - p_A - p_B - p_{AB} &= 8.2.....(iii) \\
p_0 - p_A - p_B + p_{AB} &= 2........(iv)
\end{aligned}
$$

If we add the four equations and divide by 4 we get the parameter $p_o = \frac{3.5+1.5+8.2+2}{4} = \frac{15.2}{4} = 3.8$. (i)-(ii)+(iii)-(iv) yields the value of parameter $p_A = \frac{3.5+1.5-8.2-2}{4} = \frac{8.2}{4} = 2.05$. (i)+(ii)-(iii)-(iv) yields the value of parameter $p_B = \frac{3.5-1.5-8.2-2}{4} = -\frac{5.2}{4} = -1.3$. Finally (i)-(ii)-(iii)+(iv) yields the value of parameter $p_{AB} = \frac{3.5 - 1.5 - 8.2 + 2}{4} = -\frac{4.2}{4} = -1.05$. Thus the regression equation based on the experiments is

$$y = 3.8 + 2.05x_A - 1.3x_B - 1.05x_A x_B$$

The deviation with respect to the mean is obviously given by

$$d = y - 3.8 = 2.05x_A - 1.3x_B - 1.05x_A x_B$$

It may be verified that the total sum of squares (SST) of the deviations is given by

$$\begin{aligned} SST &= 4\times\left(p^2{}_A + p^2{}_B + p^2{}_{AB}\right) = 4\times\left(2.05^2 + 1.3^2 + 1.05^2\right) \\ &= 4\times\left(4.2025 + 1.69 + 1.1025\right) = 4\times 6.995 = 27.98 \end{aligned}$$

The sample variance is thus given by

$$s_y^2 = \frac{SST}{n-1} = \frac{27.98}{3} \approx 9.33$$

Contributions to the sample variance are given by 4 times the square of the respective parameter and hence we also have

$$\begin{aligned} SSA &= 4\times 4.2025 = 16.81 \\ SSB &= 4\times 1.69 = 6.76 \\ SSAB &= 4\times 1.1025 = 4.41 \end{aligned}$$

Here SSA means the sum of squares due to variation in level of x_A and so on. The relative contributions to the sample variance are represented as percentage contributions in the following table:

	Contribution	% Contribution
SST	27.98	100
SSA	16.81	60.08
SSB	6.76	24.16
SSC	4.41	15.76

Thus the dominant factor is the machine speed followed by the depth of cut and lastly the interaction effect. In this example all these have significant effects and hence a full factorial experiment is justified.

More on full factorial design

We like to generalize the ideas described above in what follows. Extension to larger number of factors as well as larger number of levels would then be straight forward. Let the High and Low levels be represented by + an – respectively. In the case of 2^2 factorial experiment design the following will hold:

	x_A	x_B	$x_A x_B$
Row vector 1	+	+	+
Row vector 2	+	-	-
Row vector 3	-	+	-
Row vector 4	-	-	+
Column sum	0	0	0
Column sum of squares	4	4	4

We note that the product of any two columns is zero. Also the column sums are zero. Hence the three columns may be considered as vectors that form an orthogonal set. In fact while calculating the sample variance earlier these properties were used without being spelt out.

Most of the time it is not possible to conduct **that** many experiments! The question that is asked is: “Can we reduce the number of experiments and yet get an adequate representation of the relationship between the outcome of the experiment and the variation of the factors?” The answer is in general “yes”. We replace the full factorial design with a **fractional factorial design**. In the fractional factorial design only certain combinations of the levels of the factors are used to conduct the experiments. This ploy helps to reduce the number of experiments. The price to be paid is that all interactions will not be resolved.

Consider again the case of 2^2 factorial experiment. If the interaction term $x_A x_B$ is small or assumed to be small only three regression coefficients p_0, p_A, p_B will be important. We require only three experiments such that the four equations reduce to only three equations corresponding to the three experimental data that will be available. They may be either:

$$\begin{aligned} p_0 + p_A + p_B &= y_1 \\ p_0 - p_A + p_B &= y_2 \\ p_0 + p_A - p_B &= y_3 \end{aligned} \qquad (50)$$

Or

$$\begin{aligned} p_0 + p_A + p_B &= y_1 \\ p_0 - p_A + p_B &= y_2 \\ p_0 - p_A - p_B &= y_4 \end{aligned} \qquad (51)$$

depending on the choice of the row vectors in conducting the experiments.

In this simple case of two factors the economy of reducing the number of experiments by *one* may not be all that important. However it is very useful to go in for a fractional factorial design when the number of factors is large and when we expect some factors or interactions between some factors to be unimportant. Thus fractional factorial experiment design is useful when main effects dominate with interaction effects being of lower order. Also we may always do more experiments if necessitated by the observations.

One half factorial design:

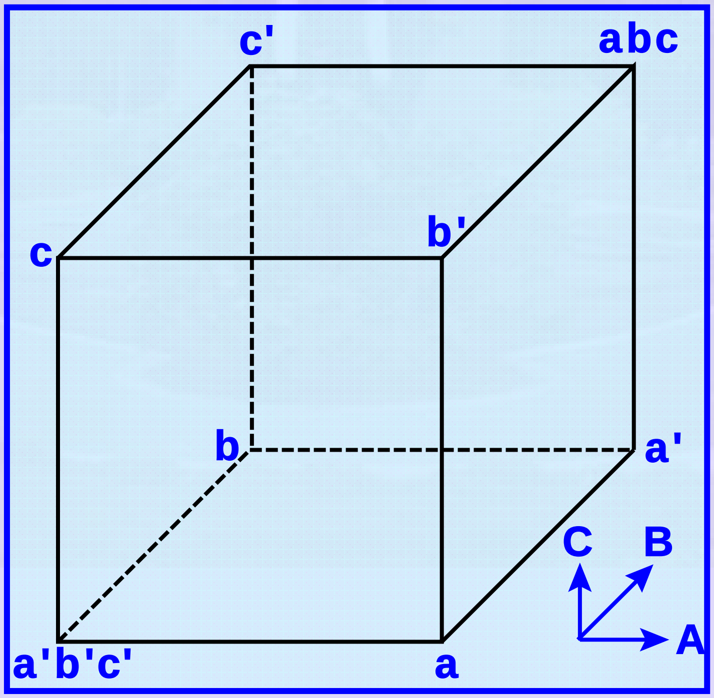

Figure 15 Three factors system with two levels. All possible values are given by the corners of the cube.

For a system with k factors and 2 levels the number of experiments in a full factorial design will be 2^k. For example, if k=3, this number works out to be $2^3=8$. The eight values of the levels would correspond to the corners of a cube as represented by Figure 15. A half factorial design would use 2^{k-1} experiments. With k=3 this works out to be $2^2=4$. The half factorial design would cut the number of experiments by half. In the half factorial design we would have to choose half the number of experiments and they should correspond to four of the eight corners of the cube. We may choose the corners corresponding to **a**, **b**, **c** and **abc** as one possible set. This set will correspond to the following:

Point	x_A	x_B	x_C
a	**+**	**-**	**-**
b	**-**	**+**	**-**
c	**-**	**-**	**+**
abc	**+**	**+**	**+**

We notice at once that the three column vectors are orthogonal. Also the points are obtained by requiring that the projection on to the left face of the cube gives a full factorial design of type 2^2. It is easily seen that we may also use the corners **a'**, **b'**, **c'** and **a'b'c'** to get a second possible half factorial design. This is represented by the following:

Point	x_A	x_B	x_C
a’	+	+	-
b’	+	-	+
c’	-	+	+
a’b’c’	-	-	-

In each of these cases we need to perform **only** 4 experiments. Let us look at the consequence of this. For this purpose we make the following table corresponding to the **a, b, c, abc** case.

Point	x_A	x_B	x_C	x_Ax_B	x_Ax_C	x_Bx_C	$x_Ax_Bx_C$
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
c	-	-	+	+	-	-	+
abc	+	+	+	+	+	+	+
Column No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Notice that Column vector 1 is identical to the column vector 6. We say that these two are **“aliases”**. Similarly the column vectors 2-5 and column vectors 3-4 form aliases. Let us look at the consequence of these.

The most general regression model that can be used to represent the outcome of the full factorial experiment would be

$$y = p_0 + p_A x_A + p_B x_B + p_C x_C + p_{AB} x_A x_B + p_{AC} x_A x_C + p_{ABC} x_A x_B x_C \qquad (52)$$

There are eight regression coefficients and the eight experiments in the full factorial design would yield all these coefficients. However we now have only four experiments and hence only four regression coefficients may be resolved. By looking at the procedure used earlier in solving the equations for the regression coefficients, it is clear that it is not possible to obtain the coefficients that form an alias pair. For example, it is not possible to resolve p_A and p_{BC}.

These two are said to be **"confounded"**. The student may verify that the following are confounded: p_B and p_{AC}, p_A and p_{BC}, p_C and p_{AB}, p_0 and p_{ABC}. The consequence of this is that the best regression we may propose is

$$y = p_0 + p_A x_A + p_B x_B + p_C x_C \qquad (53)$$

All interaction effects are unresolved and we have only the primary effects being accounted for. The **regression model** is indeed a linear model! The student may perform a similar analysis with the second possible half factorial design and come to the same conclusion.

Generalization:

Consider a 2^k full factorial design. The number of experiments will be 2^k. These experiments will resolve k main effects, ${}^kC_2$ - 2 factor interactions, ${}^kC_3$ - 3 factor interactions,…………, ${}^kC_{k-1}$ (k-1) interactions and 1 - k interactions. (if k=5, say, these will be 5 main effects, 10 – 2 factor interactions, 10 – 3 factor interactions, 5 – 4 factor interactions and 1 – 5 factor interaction).

More on simple design:

We now look at other ways of economizing on the number of experimental runs required to understand the behavior of systems. We take a simple example of characterizing the frictional pressure drop in a tube. A typical experimental set up will look like the one shown in Fig. 16.

The factors that influence the pressure drop (measured with a **differential pressure gage**) between stations 1 and 2 may be written down as:

1. Properties of the fluid flowing in the tube: **Density** ρ, the **fluid viscosity** μ (two factors)
2. Geometric parameters: **Pipe diameter D**, **Distance between the two stations L** and the **surface roughness parameter** (zero for smooth pipe and non-zero positive number for a rough pipe). (two or three factors)
3. **Fluid velocity V** (one factor)

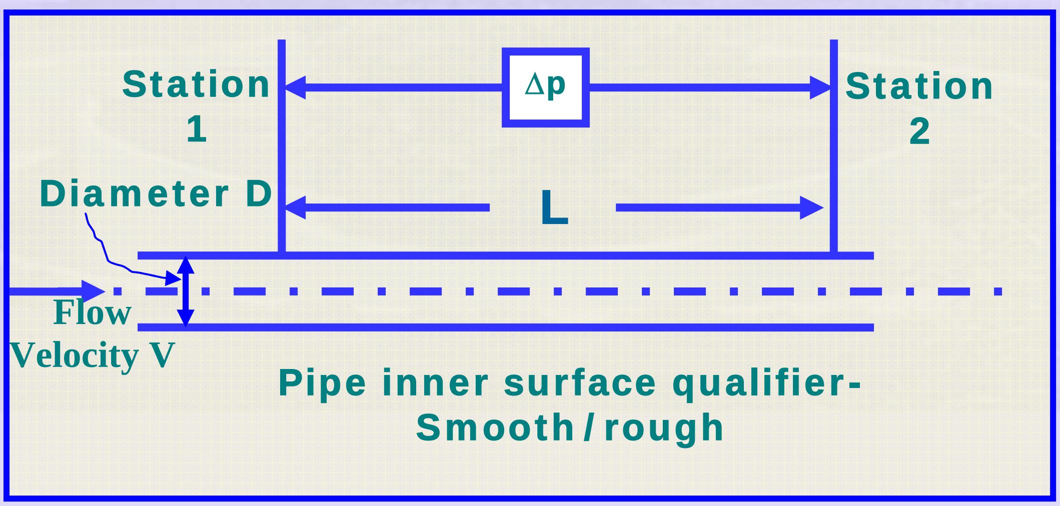

Figure 16 Schematic of an experimental set up for friction factor measurement in pipe flow

The total number of factors that may influence the pressure drop Δp is thus equal to 5 factors in the case of a smooth pipe and 6 factors in the case of a rough pipe. The student can verify that a full factorial design with two levels would require a total of **32** or **64** experiments. In practice the velocity of the fluid in the pipe may vary over a wide range of values. The fluid properties may be varied by changing the fluid or by conducting the experiment at different pressure and temperature levels! The diameter and length of the pipe may indeed vary over a very wide range. If one were to go for a simple design the number of factors and the levels are very large and it would be quite impossible to perform the required number of experiments.

The question now is: "How are we going to design the experimental scheme?" How do we conduct a **finite number of experiments** and yet get enough information to understand the outcome of the measurement? We look for the answer in **"dimensional analysis"** (the student would have learnt this from his/her course in Fluid Mechanics), for this purpose. Dimensional analysis, in fact, indicates that the **outcome** of the experiment may be represented by a simple relationship of the form

$$\boxed{\frac{\qquad}{\qquad} = \left(\frac{\qquad}{\qquad}\right)} \qquad \textbf{(54)}$$

The student may verify that both the left hand side quantity and the quantity within the bracket on the right hand side are **non-dimensional**. This means both are **pure numbers**! The f outside the bracketed term on the right hand side indicates a functional relation. In Fluid Mechanics parlance $\frac{\qquad}{\qquad}$ is known as the

Euler number and —— is known as the **Reynolds number**. The third non-dimensional parameter that appears above is the ratio L/D. Essentially the Euler number is a function of only two factors the Reynolds number and the length to diameter ratio! The number of factors has been reduced from 6 to just 2! We may conduct the experiments with just one or two fluids, a few values of the velocity and may be a couple of different diameter pipes of various lengths to identify the nature of the functional relationship indicated in Equation **(54).**

In summary:

1. *The 2k experiments will account for the intercept (or the mean) and all interaction effects and is referred to as Resolution k design. (With k=5, we have full factorial design having resolution 5. Number of experiments is 32.).*
2. *Semi or half factorial design 2^{k-1} will be resolution k-1 design. (With k=5, we have half factorial design having resolution 4. Number of experiments is 16.).*

3. *Quarter factorial design will have a resolution of k-2. (With k=5, we have quarter factorial design having resolution 3. Number of experiments is 8.). And so on…..*

The student may also work out the aliases in all these cases.