Basics of Mechanical Engineering-3

Prof. J. Ramkumar

Prof. Amandeep Singh Oberoi

Department of Mechanical Engineering

Indian Institute of Technology, Kanpur

Week 12

Lecture 53: Design of Experiments

The last lecture is on Design of Experiments. We call it DoE for short. Design of experiments is a concept that was devised by Professor Ronald Fisher in 1920. For a few decades, it was not even talked about, though mechanical engineers themselves did not discuss it much.

But when the resources to do a complete census inspection, resources to have samples in hand and then doing experiments, while doing experiments itself, some of the experiments were not even required to be done and small number of experiments would give the same kind of the output with a very significant level of maybe six sigma, it will give the results. Whatever would have been done with maybe a thousand experiments could be done in maybe 20 experiments, 30 experiments.

This was accomplished through the design of experiments. There are methods of design of experiments which I am not covering in this course, but I will give you a brief introduction. What are the various terms in the design of experiments? There are methods such as response surface methodology, factorial design, Taguchi design of experiments, and multiple ways to reduce the number of experiments while achieving similar scenario analysis.

Contents

- Design of Experiments (DoE)
- Terminology in DoE
- Steps in DoE
- Correlation
- Regression

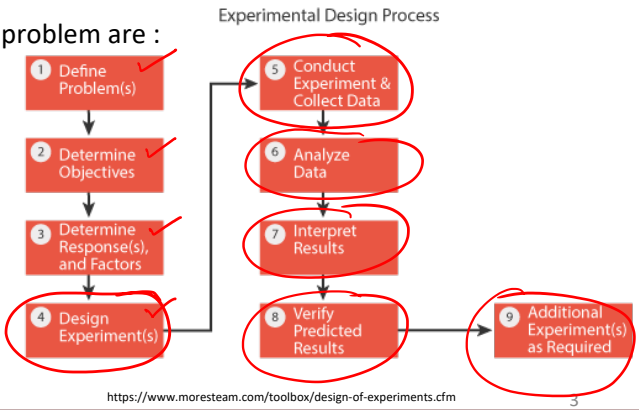
In this lecture, we will cover an introduction to design of experiments, terminologies used, steps, correlation, and regression, and how understanding regression helps formulate hypotheses for experimental design.

Design of Experiments

The purpose of statistically designed experiments is to collect appropriate data and analyse it with statistical methods resulting in valid and objective conclusions.

The two aspects of experimental problem are :

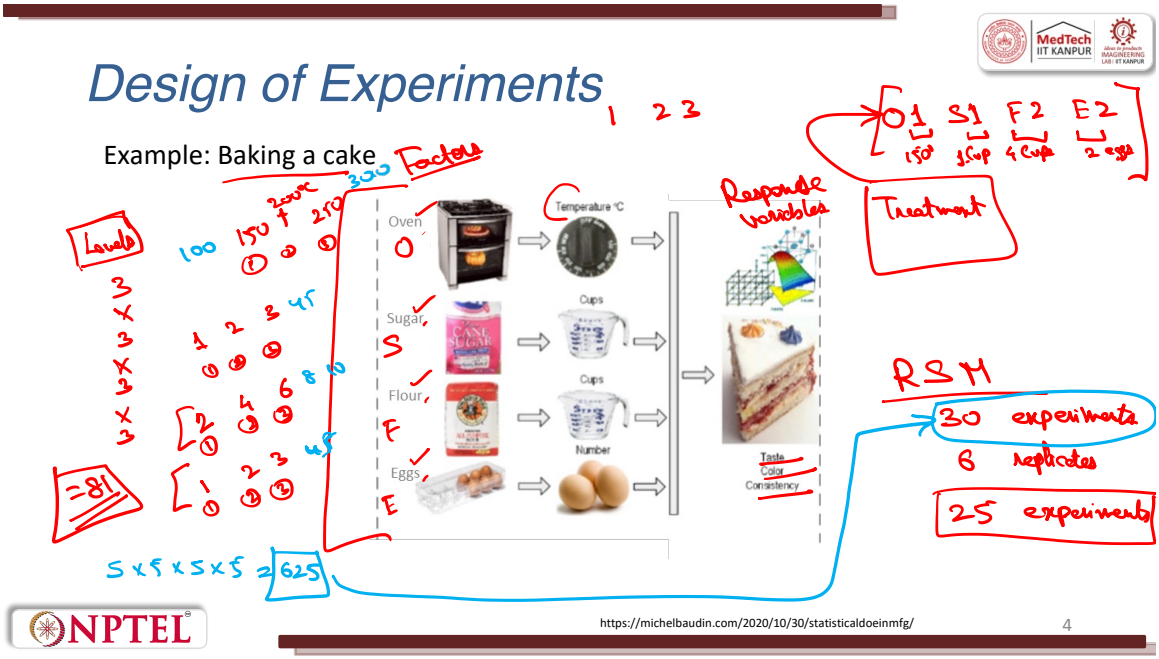
1. The design of experiment and
2. The statistical analysis



https://www.moresteam.com/toolbox/design-of-experiments.cfm

Design of experiments. The purpose of statistically designed experiments is to collect appropriate data and analyze it with statistical methods, leading to valid and objective conclusions. Two aspects of the experimental problem are the design of experiments and statistical analysis. We talked about statistical analysis. Once an experiment is designed, statistical analysis has to be conducted.

That is hypothesis testing. Whatever we discussed in the previous lecture is taken into account. The design of our experiments is important. Once we define the problem, determine the objectives, identify the responses and factors, designing experiments is then done so that we conduct them in the minimum possible time with the least expense of resources. Then we analyze the data, interpret the results, verify predicted outcomes, and conduct additional experiments for validation if required. So, what specifically is the design of experiments? I will explain.



I would like to explain the design of experiments concept with this example. For instance, in machining, you must achieve a specific level of roughness. For that roughness in machining, certain parameters are selected.

For example, cutting speed (spindle rotation), depth of cut, feed rate, and the type of tool used. These become the parameters to be finalized or optimized to achieve the desired roughness for your application where the manufactured component will be used. Now, let

me take the example of baking a cake. For baking a cake, certain parameters are involved. For example, oven temperature, sugar, flour, and eggs.

These are the resources available. And, for instance, if you have three levels of each—the temperature, for example, varies from 150 degrees to 250 degrees—the center point could be 200 degrees. I am talking about degrees centigrade. Then, sugar could be 1 cup, 2 cups, or 3 cups. Flour could be, say, 2 cups, 4 cups, or 6 cups.

Eggs could be 1 egg, 2 eggs, or 3 eggs. So, these three—1, 2, 3—here also 1, 2, 3, totaling three. Three each becomes levels. I'll talk about the terminology. The total number of experiments needed to find a perfect combination—or the ideal result—of the right balance between oven temperature, sugar amount, flour, and eggs.

We would have to conduct three experiments multiplied by 3, by 3, by 3, equaling 81 experiments. Conducting 81 experiments—and possibly ruining some cakes due to extreme conditions we don't even need—is not a great idea. That's why the design of experiments is required. Here, what are we doing? We are trying to adjust the temperature, perhaps to one specific level. For example, let's say parameters O, S, F, E, each having levels 1, 2, and 3. I will say O1, S1, F2, and E2.

What does this mean? O1 is set at 150 degrees centigrade. This is 1 cup of sugar. This is F2; flour 2 means 4 cups of flour. This 2 is 4 for the flour.

And E2 means 2 eggs. This one specific set that we are working on is known as one treatment. Like this, I talked about how we would have to do 81 treatments, and out of these 81 treatments, after baking the cake, what do we need to identify? We would have response parameters here. I will call the response variables, which are taste, color, and consistency.

The taste of the cake with more sugar would be more tangy. With higher temperature, the color would become darker. The consistency would depend on maybe eggs or some other factors. So, we do not know which of these, when I say factors, I am talking about the input variables for which different levels are there. In case of baking 81 cakes, in place of 81 experiments, there could be half factorial in which half of the experiments would be conducted.

And in factorial analysis or in response surface methodology, generally the extremities are always taken. For example, 1, 1, 1, 1 level of all four factors here and 3, 3, 3, 3 level of all four factors here, and the middle level is also taken. Then, combinations like 1, 2, 1,

3, or similar, are decided based on the kind of design of experiments we are working on. So, this is what we would have conducted with four variables. Methodologically, as I mentioned, RSM with four factors would suggest 30 experiments at most.

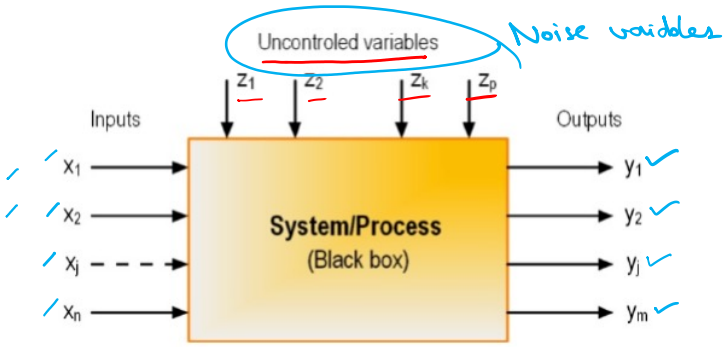
In these 30 experiments, we also have six replicates. Why do I say replicates? Because if one time you have taken the best cake combination—for example, at specifically level 1, 2, 3, 1—you may have gotten a specific combination that produces the best or most acceptable cake, which you will finally present to the customer with the best taste, color, and consistency. That is only one time you have done it. Replication means this experiment would have to be repeated at least three times to see whether the results are replicable or not.

Here, out of 36, some are replicates. If I remove the replicates, 25 experiments are sufficient instead of 81. Now, for instance, you might have 5 levels: 100 degrees, 150 degrees, 200 degrees, and 300 degrees. Then, we will have 1, 2, 3, 4, and 5 cups of sugar; 2, 4, 6, 8, and 10 cups of flour; and 1, 2, 3, 4, and 5 eggs. Here, what will be the total number of experiments?

That would be 5 into 5 into 5 into 5, which is equal to 625 experiments. A very large number of experiments, but still, RSM will suggest only 30 experiments. So, this is the magic of the design of experiments. You understand all the extremities. Then you try to see what the center point is and how you observe the variation in between, so you can have experiments within the design.

Design of Experiments

Example:



Now, what do we have here? In any kind of setup, this is the baking of a cake. It could have been selecting the parameters for running a turbine. For running a turbine, there are four parameters selected. For example, four parameters each have their settings, and each setting is designed for a specific condition at which the turbine would have maximum efficiency.

Or maybe we will say where in the turbine blades there should be minimum efficiency. Two or three responses we can select. In the case of machining itself, the response parameters could be quality and also performance. Performance, in a way; quality would be the surface roughness we are getting. The performance would be what the tool wear is.

Tool wear is to be minimized. Surface roughness is to be set to a specific level. So, the way we design it depends on those output parameters: y1, y2, till ym, varying from yj. These are the output parameters we are working on. So, we aim—or the design of experiments aims—to correlate these outputs: y1, y2, and so on, whether it's taste consistency, color, surface roughness, tool wear, or cutting fluid wastage. Based on the input parameters—such as speed, feed, depth of cut, the kind of cutting fluid used, the kind of tool material used, or the material-tool combination used—all this happens in the presence of uncontrolled variables. When I say uncontrolled variables, I mean z1, z2, and so on up to zc, which are also known as noise variables.

For example, in this cake, I considered only four factors: oven, sugar, flour, and eggs. What about additives? What about the quality of sugar? Are we using fine-grained sugar, or are we getting sugar from a different store? What about the kind of egg we are using?

What about other parameters? The way you mix, the time you mix for it. So, these are all parameters you are fixing, which you are not even varying. Only the most significant parameters are being taken into account. Other parameters could also be there, known as uncontrolled variables. In machining itself, uncontrolled variables exist even when the machine should be perfect.

There could be a machine which is maybe 20-25 years old and has been used. What about the machine's performance? What about the human who is working on the machine? Those are known as uncontrolled variables. So, this is how the design of an experiment is to be conducted, and we try to have the overall output.

Terminology in DoE

Terms used in DoE

- **Factor**: A variable or attribute which influences or is suspected of influencing the characteristic being investigated.
- All input variables which affect the output of a system are factors. They can be qualitative (type of material, type of tool, etc.) or quantitative (temperature, pressure, etc.).

- **Levels of a factor:** The values of a factor variable being examined in an experiment. If the factor is an attribute, each of its state is a level. For example, setting of a switch on or off are the two levels of the factor switch setting. If the factor is a variable, the range is divided into required number of levels. The levels can be fixed or random.

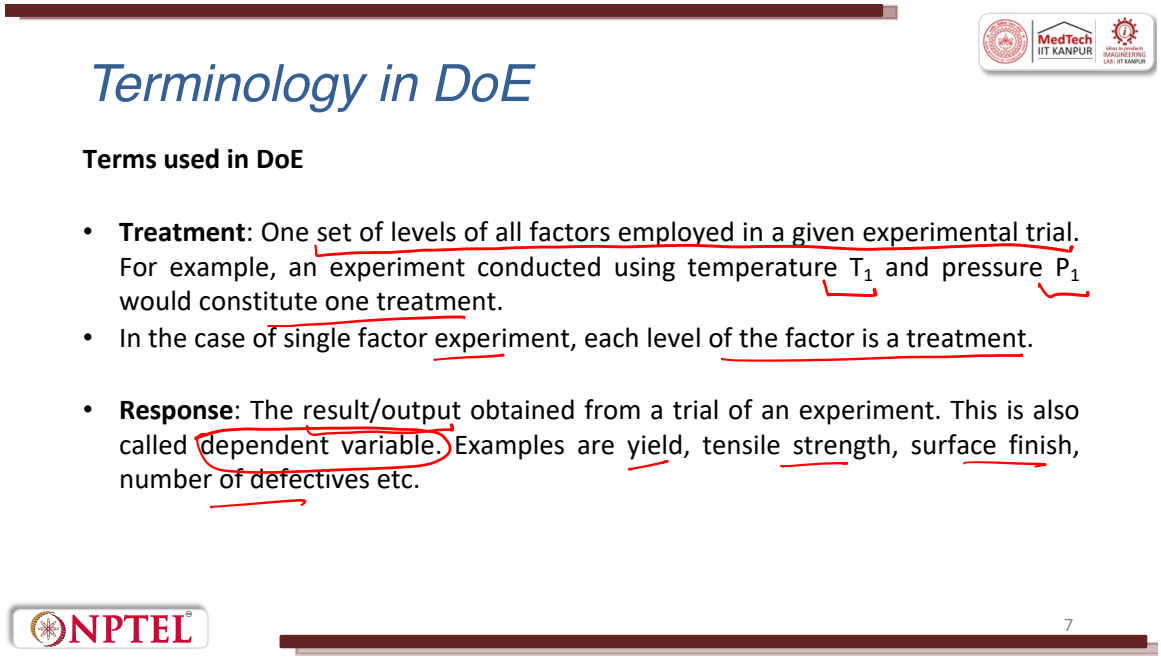
Now, let me try to talk about the terms used in the design of experiment factors. This is a variable or attribute which influences or is suspected to influence the characteristic being investigated. All input variables affect the output of a system or factors.

They can be qualitative, that is, the type of material, type of tool, or quantitative, that is, temperature, pressure, etc. We have talked about quantitative and qualitative.

Quantitative is metric, again to recall, and this is non-metric. Levels of a factor, levels I talked about. Up to three levels I worked for, three x, up to five levels I worked for.

Levels of a factor, the values of a factor, variable being examined in an experiment. If a factor is an attribute, each of its states is a level. For example, the setting of a switch, on and off, are two levels of a factor of switch setting. If the factor is a variable, the range is divided into the required number of levels. The levels can be fixed or random.

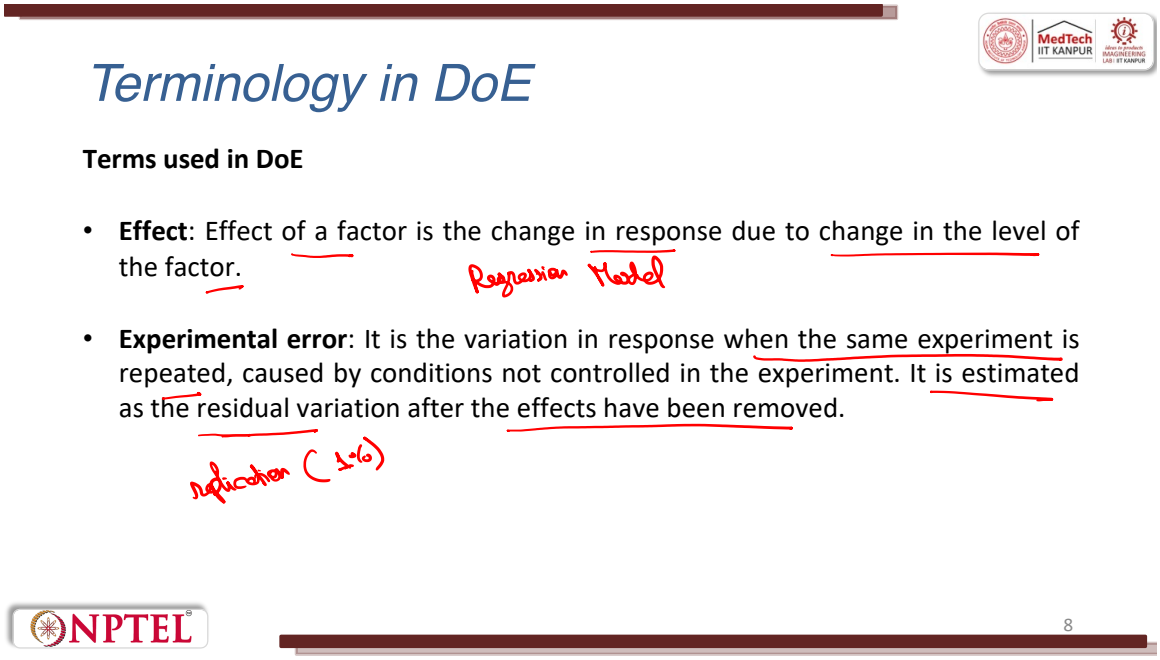
It could be fixed or random. For instance, the temperature: I took only 100, 150, 200, 250, and 305 levels. So, I fixed this specific range. It could have been 201 degrees, 202 degrees, but I fixed it to a specific level. What is happening between 150 and 100 would be given by the curve plot that we get through.



Then comes treatment. Treatment, as I said, is one set of levels of all factors employed in a given experimental trial. For example, an experiment conducted using temperature T1 and pressure T1 would constitute one treatment. What I said here was, this is one treatment. O1 at a specific level, 1.

Sugar, level 1. Flour, level 2. Eggs, level 2. This is one treatment. In a single-factor experiment, each level of the factor is a treatment.

Response. The result or output obtained from a trial of an experiment. This is also called the dependent variable. These factors are also known as independent variables. Examples are yield, tensile strength, surface finish, number of defects, etc.



Then comes effect. The effect of a factor is the change in the response due to a change in the level of the factor. This effect is generally presented through a regression model. Experimental error. It is the variation in response when the same experiment is repeated.

I talk about replication. Replication, within what range the replication is happening. For example, if I measure the roughness, each time the roughness comes to 2.5 micrometres. But it varies between 2.51 and 2.53 micrometres. That is the only upper level of tolerance already coming.

Taking a percentage of 0.02 by 2.5, maybe I could say within 1% of control it is happening. So, this error is estimated as the residual variation after the effects have been removed.

Terminology in DoE

Designing of experiment requires attention to four potential traps that can create experimental difficulties:

1. In addition to experimental error, other sources of error or **unexplained variation**, can confuse the results.
2. Uncontrollable factors that induce variation under normal operating conditions, referred to as "**Noise Factors**".
3. **Correlation** can often be confused with causation.
   Two factors, varying together, may be highly correlated while one doesn't cause the other - they may both be having been caused by a third factor.
4. The **combined effects** between factors require careful planning prior to conducting the experiment.

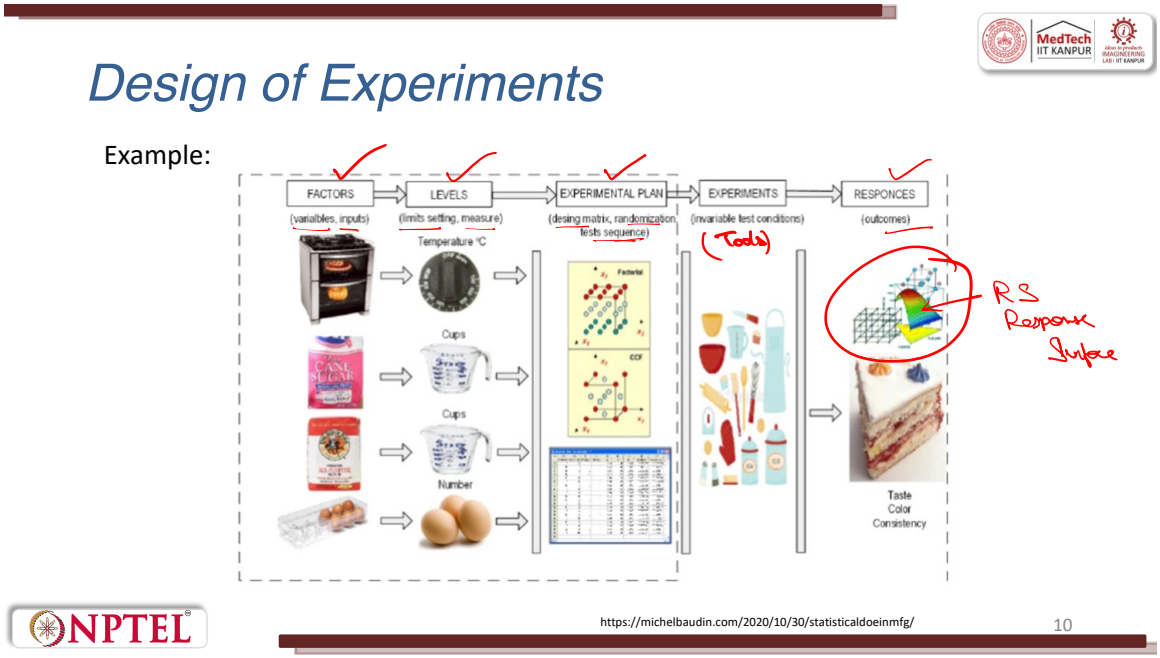
Designing an experiment requires attention to four potential traps that can create experimental difficulties. In addition to experimental error, other sources of error, that is, unexplained variation, can confuse the results. Unexplained variation: the person conducting the experiments is not serious.

The responses are not being observed through a good instrument. Unexplained variance could also be there. Uncontrollable factors that include variation under normal operating conditions referred to as noise factors, these are just mentioned about. Correlation can often be confused with causation. Correlation only gives, this is going up and down. Cause is never there in correlation.

Correlation is only talking about, this is happening when other is happening; this is increasing or decreasing, we'll talk about. Correlation causation is a different scenario. Two factors varying together may be highly correlated, but one doesn't cause the other; they may both have been caused by a third factor. We do not know.

This is to be measured. Combined effects between factors require careful planning before conducting the experiment. When I say combined effect, in case of, if you vary two variables for example number of eggs and number of cups of sugar, you increase simultaneously, there would be a combined effect of both of them. We do not know what has enhanced or reduced the level of consistency in the cake you bake. Or if you increase the feed and decrease the speed together, we do not know what has contributed to the change in roughness.

For that, again, a number of experiments are to be conducted, and these are to be very carefully planned. And all this is based upon previous experience and the secondary data which people have provided where they have conducted similar experiments. Accordingly, there always has to be a starting point. The starting point is taken through one factor at a time, keeping everything else constant. Then we see what the range of the factors is that we will try to work upon for experimentation.

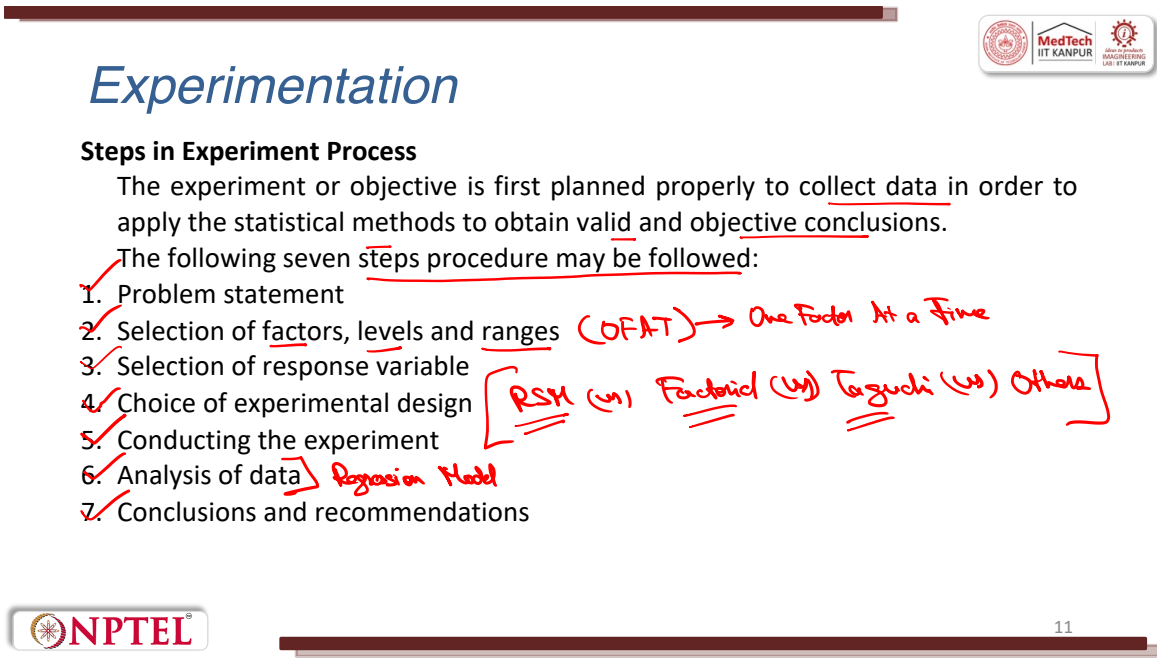


Now, this is a more detailed version of the illustration which I showed you. I only showed you factors and levels in the previous part. Factors are also known as variables or inputs. Levels are the limit settings or measures. And I showed you responses; that's the outcomes. Outcomes would come in these kinds of graphs.

This kind of graph that you see is a response surface. In between, we have the experimental plan. That is the design matrix, randomization, test sequence—which test we conducted first or later. Then, doing experiments. Actually, for experiments, we need to have all the tools ready with us.

It is invariable test conditions. The person who is conducting the experiments. The machine should stay consistent. If you are using one oven, for example, one make oven, we are using Murphy Richards oven. Let us keep doing it in Murphy Richards only.

We should not move to Philip or another set of experiments within the same set of the decision that you are trying to make. So that means there should be consistency steps in experiment process.



The experiment or objective is first planned properly to collect data in order to apply statistical methods to obtain valid and objective conclusions. The following seven-step procedure may be followed. Define the problem statement.

Select factors, levels, and the ranges of these factors. This is done by OFAT. I am giving you terms. You can refer to the books given in the reference slide here to read more about them. This is one factor at a time.

Selection of response variables. Which response variables are more required to be studied? Choice of experimental design, whether we are using RSM versus factorial versus Taguchi method versus others. What kind of experimental design are you trying to have here? Conducting the experiments, then analyzing the data, drawing conclusions, and making recommendations.

Experimentation

Improving process or product "Robustness"

- **Robust design:** It is an engineering approach to create products and processes that consistently perform well despite variations in manufacturing, usage or environmental conditions. The system should be insensitive to "noise factors" (uncontrollable variations), or factors cause minimal impact, leading to improved reliability and reduced costs associated with rework and defects.
- **Balancing Tradeoffs:** It refers to a decision-making process considering advantages and disadvantages of different options, often involving give-up something desirable to gain something else (Trade-off one thing for another).
  To balance tradeoffs, one has to identify the options, evaluate pros and cons of them and select the option that can give the best overall outcome, making a compromise to lose some other option which may give few much better but overall lesser outcome.

In experimentation, improving the process or product robustness is most important. Robust design is an engineering approach to create products and processes that consistently perform well despite variations in manufacturing, usage, or environmental conditions. The system should be insensitive to the noise factors, which are uncontrollable variations or factors that cause minimal impact, leading to improved reliability and reduced costs associated with rework and defects. Balancing trade-offs. A trade-off is between one thing and another.

It is a decision-making process considering the advantages and disadvantages of different options. Often, it involves giving up something desirable to gain something else. Trade one thing for another. There will always be a trade-off. You might find the cake that is best with one set of settings.

There could be another set of settings where you also find the cake that is also very great. But now, which one of them is using the minimum resources? If the oven temperature is higher, the energy consumption is higher. If the number of eggs is more, the cost would also go up. Then there has to be the right selection of the optimal results that are there, so you have that with the minimum use of resources, number one.

Number two, there could be a trade-off. With the same cost, with the same energy, you have two sets. In one, the number of eggs is two; in the other, the number of cups of sugar is a little more, maybe one cup more than the previous one. Two sets are completely ready. Then either you use more cups or you use more eggs.

There could be a trade-off. One could have been given up for the other. Or, in other words, I would like to—I'm talking now only about factors or variables, input variables. In the response variables, the trade-offs are more important to be decided upon. Maybe, for example, between the consistency and color.

When the taste is all equal, what would you select? Consistency or color? People sometimes look at the cake, buy, purchase the cake because of the color itself. When they taste them, they will like consistency more. So, one thing have to be completely washed away.

For example, you completely wash away. I do not now prefer color consistency has to be there. It may be in case of tool where we see roughness, the tool where is going little higher, but reference is better. You say okay, I will try to give away. One thing for another, that is trade-off, one thing for another to gain something, I will have to lose something.

This is trade-off, to balance trade-off. To balance trade-off, one has to identify the options, evaluate pros and cons of them and select the option that can give the best overall outcome, making a compromise to lose some option than other which may give few much better but overall lesser outcome.

So, I am not talking about these methods here, responsible methodology, factorial method, Taguchi method or any other. This you can read. Owing to the time availability, I have given you brief introduction that design of experiment is required and that is one of the important or critical thing when you try to really conduct the experiments with the available or minimum resources.

Correlation

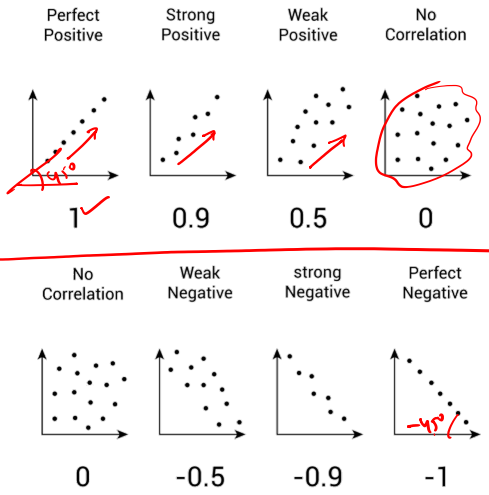
- **Correlation** is a statistical measure to describe the extent to which two variables are related to each other.
- It indicates whether they tend to change together in a consistent way.
- This relationship can be plotted and displayed on a scatter plot and is quantified by a correlation coefficient, which ranges from -1 to 1.
  - A positive correlation means variables increase or decrease together.
  - A negative correlation means they move in opposite directions.
  - A correlation near zero indicates no linear relationship.
- Correlation does not imply causation, meaning just because two variables are related does not mean one directly causes the other.

Now, let me talk about correlation and regression. When you conduct the design of experiments, in the steps you could see there is analysis of data. This analysis of zeta as I mentioned, generally we try to have a regression model.

To have a regression model, we need to understand what correlation is and what regression is. Correlation is a statistical measure that describes the extent to which two variables are related to each other. It indicates whether they tend to change together in a consistent way. This relationship can be plotted and displayed on a scatter plot and is quantified by a correlation coefficient, which ranges from minus 1 to plus 1. A positive correlation means variables increase or decrease together.

A negative correlation means variables move in opposite directions. A correlation near 0 indicates no linear relationship. Correlation does not imply causation. This is always true. Causation and correlation should not be confused. Meaning, just because two variables relate to each other does not mean one directly causes the other.

Correlation



https://www.simplypsychology.org/wp-content/uploads/correlation-coefficient.jpg

So here is a plot. When I say a correlation coefficient value of one, it is a perfect positive correlation with a change in one. The other thing is directly changing. For instance, by adding sugar, the sweetness would increase. So, this is a positive correlation.

Strong positive, which is not perfectly one, but yes, it is also varying in a positive direction. Weak positive, it has some scatter, but it is still a positive correlation. Zero, a completely scattered plot with no correlation. Exactly the opposite direction, this is a 45-degree angle, and when we have a negative 45-degree angle, this is a minus 1 correlation, 0.9, 0.5, 0, no correlation. Now, let me talk about regression.

Regression

Regression: It is a method for studying the relationship between variables, particularly how a dependent variable changes in response to one or more independent variables. It aims to:

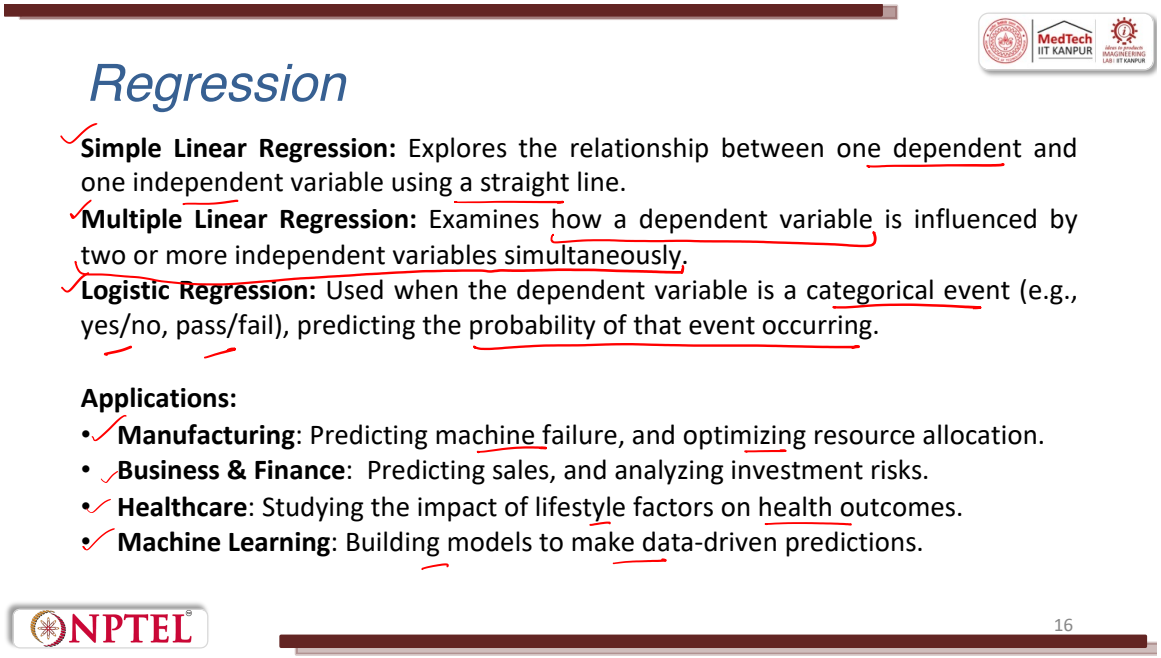
Find patterns and trends in data.
Quantify the strength and direction of the relationship between variables.
Predict about dependent variable based on changes in independent variables.

Key Concepts:

- **Dependent Variable (Y)**: The outcome or response variable to be predicted.
- **Independent Variable(s) (X):** The input variables that are believed to influence the dependent variable.
- **Regression Model:** A statistical model to establish a mathematical equation to describe the relationship between the (X) and (Y), often represented by a line or curve that best fits the data points.

Regression is a method for studying the relationship between variables, particularly how a dependent variable changes in response to one or more independent variables. It aims to find patterns and trends in data, quantify the strength and direction of the relationships between variables, and predict the dependent variable based on changes in the independent variable.

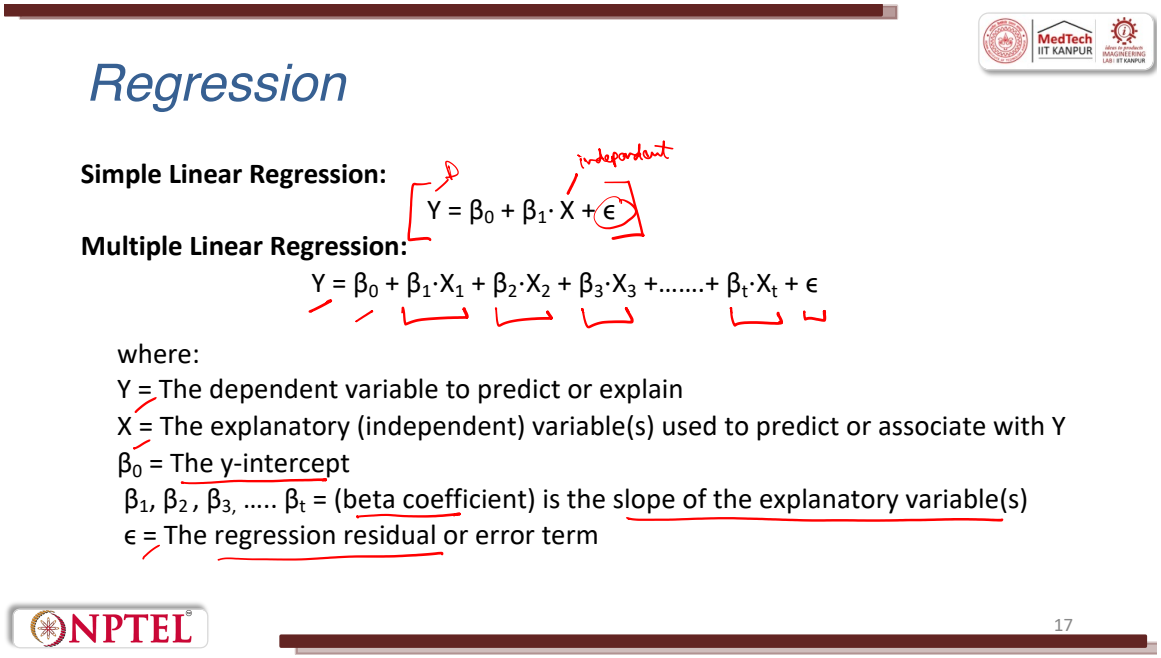
Key concepts: There is a dependent variable, which I call the response variable. This is the outcome or response variable to be predicted. There is an independent variable, which I call factors, which are input variables that I believe influence the dependent variable. This is a regression model. That is a statistical model to establish a mathematical equation describing the relationship between X and Y, often represented by a line or a curve that best fits the data points.



Simple linear regression, multiple linear regression, and logistic regression are three types. There could also be nonlinear regression as well. Simple linear regression explores the relationship between one dependent and one independent variable using a straight line. Multiple linear regression examines how a dependent variable is influenced by two or more independent variables simultaneously. The examples I took, baking a cake, machining, those were all multiple linear regression examples. We are trying to identify

or see how surface roughness depends on the speed, the feed, the depth of cut, and the type of tool.

So, that could be multiple regression. Logistic regression is something different. This is used when the dependent variable is categorical. For example, yes, no, pass, fail—predicting the probability of that event occurring. Applications are in manufacturing, such as predicting machine failure; optimizing resource allocation in business and finance; predicting sales; analyzing investment risk in healthcare; and studying the impact of lifestyle factors on health outcomes. Machine learning involves building models to make data-driven predictions.



Simple linear regression is expressed this way: $Y = \beta_0 + \beta_1 \cdot X + \epsilon$. This is the dependent variable. This is the independent variable. Epsilon is the regression error term, meaning the model we have developed is not a perfect match. There is an error. This error must be minimized. There will be multiple models derived from the analysis.

Whichever model has the minimum error will be the closest. Multiple linear regression:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \ldots\ldots + \beta_t \cdot X_t + \epsilon$$

where:

Y = The dependent variable to predict or explain

X = The explanatory (independent) variable(s) used to predict or associate with Y

β_0 = The y-intercept

$\beta_1, \beta_2, \beta_3, \ldots. \beta_t$ = (beta coefficient) is the slope of the explanatory variable(s)

ϵ = The regression residual or error term

Regression

Multiple Linear Regression:
Example: Prediction of Gas Turbine Performance.

In a combined-cycle power plant utilizing gas turbines and a steam turbine, the **net electrical power production** (the dependent variable) can be represented as a linear function of many independent variables, including:

AT: Ambient Temperature (in °C)
V: Exhaust Vacuum (in mmHg) - pertaining to the steam turbine
AP: Ambient Pressure (in millibars)
RH: Relative Humidity (dimensionless quantity)

Let me see an example. Multiple linear regression for prediction of gas turbine performance. Gas turbine performance is how the turbine would run; it will produce electricity. Production of electricity, that is power output, is dependent upon what variables? In a combined cycle power plant, which is also known as CCPP, utilizing gas turbines and a steam turbine.

Both of them together, the net electrical power production, that is the dependent variable, can be presented as a linear function of many independent variables, including the ambient temperature at 80 degrees centigrade, exhaust vacuum in millimeters of mercury (that is pertaining to the steam turbine), ambient pressure in millibars, and relative humidity, which is a dimensionless quantity.

Regression

Multiple Linear Regression:

The equation for multiple linear regression is expressed as:

$$PO = \beta_0 + \beta_1 \cdot AT + \beta_2 \cdot V + \beta_3 \cdot P + \beta_4 \cdot RH + \epsilon$$

Where:

- PO denotes the anticipated Electrical Power Output.
- β_0 is the y-intercept.
- β_1, β_2 , β_3, and β_4 are the coefficients for each variable, indicating the change in PO for a one-unit alteration in the respective variable, providing all other variables are held constant.
- ϵ represents the error term.

This is a multiple regression model for the turbine power output. Here, $PO = \beta_0 + \beta_1 \cdot AT + \beta_2 \cdot V + \beta_3 \cdot P + \beta_4 \cdot RH + \epsilon$. Now, here β_0—you know what it is?

Where:

- PO denotes the anticipated Electrical Power Output.
- β_0 is the y-intercept.
- β_1, β_2 , β_3, and β_4 are the coefficients for each variable, indicating the change in PO for a one-unit alteration in the respective variable, providing all other variables are held constant.
- ϵ represents the error term.

Regression

Multiple Linear Regression:

Utilization in Turbines
Engineers utilize past operational data obtained from SCADA systems to compute the values of the β coefficients.

The resultant model enables them to:

1. Forecast Power Output: Anticipate the power generation of the turbine under varying climatic conditions, such as a hot, humid day compared to a cold, dry day.
2. Performance Evaluation: Establish a benchmark for anticipated optimal performance.

A substantial discrepancy between the actual power production and the model's projection may indicate performance deterioration or a possible malfunction in the turbine components.

Now, utilization turbines is one equation taken from the University of California database, where they have given this combined cycle power plant system. Utilization turbine engineers utilize past operational data obtained from the SCADA system to compute the values of beta coefficients.

The resultant model often enables them to forecast power output—that is, to anticipate the power generation of the turbine under varying climatic conditions, such as a hot, humid day compared to a cold, dry day. Performance evaluation establishes a benchmark for anticipated optimal performance. A substantial discrepancy between the actual power production and the modest projection may indicate performance deterioration or a possible malfunction in the turbine components.

Regression

3. Analysis of Efficiency: Examine the coefficients to measure the influence of each variable on overall efficiency.

A markedly negative β for Ambient Temperature (β_1) corroborates the established technical premise that gas turbine performance diminishes substantially with increasing ambient temperature.

Then, analysis of efficiencies also helps them. Examine the coefficients to measure the influence of each variable on overall efficiency. A marked negative beta for ambient temperature corobots established technical premise that gas turbine performance diminishes substantially with the increasing ambient temperature. So, let me show you.

Regression

The estimated beta values for a linear model are often close to the following (based upon combined cycle power plant (CCPP) dataset:

$$PO \approx 454.61 - 1.98 \cdot AT - 0.23 \cdot V + 0.06 \cdot AP - 0.16 \cdot RH$$

Intercept	$\beta_1 \approx 454.61$	Baseline predicted power output (in MW, for example).
AT	$\beta_1 \approx -1.98$	Power output decreases by about 1.98 MW for every 1 ∘ C increase in Ambient Temperature.
V	$\beta_2 \approx -0.23$	Power output decreases by about 0.23 MW for every 1 mmHg increase in Exhaust Vacuum.
AP	$\beta_3 \approx 0.06$	Power output increases by about 0.06 MW for every 1 millibar increase in Ambient Pressure.
RH	$\beta_4 \approx -0.16$	Power output decreases by about 0.16 MW for every 1-unit increase in Relative Humidity (if RH is on a 0-100 scale, this is per 1% point).

https://archive.ics.uci.edu/dataset/294/combined+cycle+power+plant

The estimated beta values for a linear model are often close to the following (based upon combined cycle power plant (CCPP) dataset:

$$PO \approx 454.61 - 1.98 \cdot AT - 0.23 \cdot V + 0.06 \cdot AP - 0.16 \cdot RH$$

Intercept	$\beta_1 \approx 454.61$	Baseline predicted power output (in MW, for example).
AT	$\beta_1 \approx -1.98$	Power output decreases by about 1.98 MW for every 1 ∘ C increase in Ambient Temperature.
V	$\beta_2 \approx -0.23$	Power output decreases by about 0.23 MW for every 1 mmHg increase in Exhaust Vacuum.
AP	$\beta_3 \approx 0.06$	Power output increases by about 0.06 MW for every 1 millibar increase in Ambient Pressure.
RH	$\beta_4 \approx -0.16$	Power output decreases by about 0.16 MW for every 1-unit increase in Relative Humidity (if RH is on a 0-100 scale, this is per 1% point).

These are known as interaction effects. For this, you can read the book that I have given. There is a book by Douglas Montgomery, which I have given in the references, and there are multiple other books which are given in the references. You can read about them. If you have any queries, you can come up to the forum for any questions.

To Recapitulate

- What is the purpose of DoE?
- Define factors and levels in DoE.What are the key steps in DoE?
- Why is randomization important in experiments?
- What does correlation coefficient measure?
- How do we interpret positive correlation?
- What is regression analysis used for?
- Difference between simple and multiple regression?

Just to recapitulate what we discussed in this lecture, we talked about the purpose of the design of experiments. We saw the terminology: factors, levels, treatments. What are factors and levels in DOE? This is one of the questions.

What are the key steps in the design of experiments? Why is randomization important in experiments? What does the correlation coefficient measure, and how do we interpret a positive correlation? What is regression analysis used for? Difference between simple and multiple regression.

With this, I'm concluding my talk and lecture series on the design of experiments and also on statistics in mechanical engineering. With this, I'm also concluding the Basics of Mechanical Engineering three-course, which is a 12-week course. We discussed thermodynamics and fluid mechanics.

This also concludes the overall series of the Basics of Mechanical Engineering by Professor Ramkumar and Dr. Amandeep Singh, that is, me. We hope that with the virtual demonstrations, tutorial sessions, and theory, you have learned. If you have any queries, we are open to any input and feedback you provide.

References:

1. Madsen, B., 2016. Statistics for non-statisticians. Springer.
2. Wolverton, M.L., 2009. Research Design, Hypothesis Testing, and Sampling. Appraisal journal, 77(4).
3. Berryman, S., 2020. the Mechanical Hypothesis. The Cambridge Companion to Ancient Greek and Roman Science, p.229.
4. Montgomery, D.C., 2017. Design and analysis of experiments. John wiley & sons.
5. Krishnaiah, K. and Shahabudeen, P., 2012. Applied design of experiments and Taguchi methods. PHI Learning Pvt. Ltd..

Thank you.