

**Statistics for Experimentalists**  
**Prof. Kannan. A**  
**Department of Chemical Engineering**  
**Indian Institute of Technology - Madras**

**Lecture - 37**  
**Regression Analysis: Part A**

Welcome back to the course on statistics for experimentalists. Today, we will be talking about regression analysis. So far we have been looking at design of experiments, factorial design, fractional factorial design. Will take a small break from the design of experiments and look at regression analysis. You will also find a lot of similarities between design of experiments and regression analysis.

For example, the analysis of variance concept will also be extensively used. The t-test will also be used in the regression analysis. Simply put regression analysis indicates development of empirical correlations to the experimental data. We are not modeling from first principles. We are trying to find the relationship between the factors that are influencing the experiment and the response recorded from the experiments.

This has lot of significance, it is not giving the data to a spreadsheet or to a curve fitting program and getting a high value of R squared. You all know what is R squared and we aim for R squared values of 0.99 and we add as many terms as possible to remodel equation to achieve this high R squared. The important thing is the models should be simple and it should not be unwieldable.


So the more number of terms you have in the model equation even though it may look impressive on paper. It will be very difficult to apply and the predictive capabilities of the model may also decline. It may work well for a given set of data but may not do so for other set of data from some other source. So anyway with this brief introduction let us get started.


Rather than looking at simple least squares method, we will be applying linear algebra concepts to do multiple regression analysis.

**(Refer Slide Time: 02:37)**

## References

Draper, N. R., H. Smith, *Applied Regression Analysis*.  
3<sup>rd</sup> ed. New Delhi: Wiley-India, 1998.

Montgomery, D. C., G.C. Runger, *Applied Statistics  
and Probability for Engineers*. 5th ed. New Delhi:  
Wiley-India, 2011.





The references for the subject are Draper and Smith Applied Regression Analysis, third edition, New Delhi, Wiley India and the book by Montgomery and Runger, Applied Statistics and Probability for Engineers, fifth edition, New Delhi, Wiley.

**(Refer Slide Time: 02:54)**

## References

Montgomery, D. C., *Design and Analysis of Experiments*.  
8th ed. New Delhi: Wiley-India, 2011.

Kutner, M. H., C. J. Nachtschiem, J. Netner, *Applied  
Linear Regression Models*. 4<sup>th</sup> ed. New Delhi: McGraw  
2004.



And this is the prescribed text book for the course on statistics for experimentalists. You also have the book written by Montgomery, Design and Analysis of Experiments and another good book is written by Kutner, Nachtschiem, Netner, Applied Linear Regression Models, fourth edition, New Delhi published by McGraw Hill.

**(Refer Slide Time: 03:24)**


## Multiple Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Y : Response

$X_1$ : independent variable 1

$X_2$ : independent variable 2

 independent error term

So now let us come to the multiple regression model, so we will take simple model first where Y is the response and the model parameters are beta 0, beta 1, beta 2, X1 and X2 are the factors or the independent variables, epsilon is the error term. This is a very interesting model. We say that the response Y is governed by a combination of 2 factors X1 and X2. The important question is whether this error term is because of random effects alone.

We have started cautiously with a simple model and then the unaccounted extra effects are ascribed to this error term. The error may be random or may be a combination of random effects and also the unexplained portion from the model. If this model was inadequate, then the error term will absorb the unaccounted part of the response. So in such situations, the epsilon term here cannot be thought of random error.

**(Refer Slide Time: 05:10)**

## Multiple Regression Model

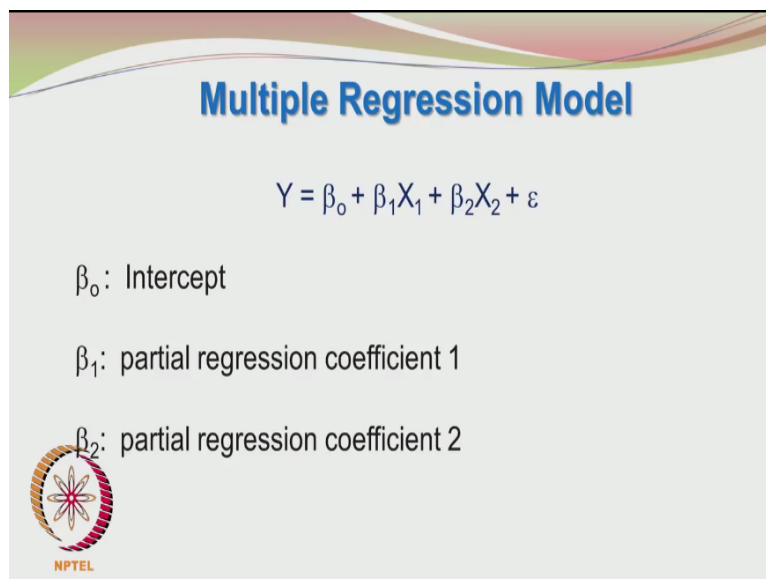
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

The term multiple appears because it contains more than one regressor variable and the term linear appears because the equation is a linear function of the unknown parameters  $\beta_0, \beta_1$  and  $\beta_2$ .

Now we are talking about multiple regression models and why is it called multiple regression.  $X_1$  and  $X_2$  are called as regressor variables and if there is more than one regressor variable, we use the term multiple and the term linear is usually associated with these kind of regression models because the parameters  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  are linear in nature.

And here if you want to make it nonlinear, you can develop a model  $Y = \sin \beta_1 X_1$  or  $Y = \beta_0 + e^{\beta_1 X_1}$ . Then you cannot term the regression model as linear because the parameters are nonlinear in nature.

**(Refer Slide Time: 06:39)**




**Multiple Regression Model**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$\beta_0$ : Intercept

$\beta_1$ : partial regression coefficient 1

$\beta_2$ : partial regression coefficient 2

 NPTEL

Before we proceed further into the course, I would like to recommend something, please refresh your concepts on linear algebra. You do not have to go very deep into the subject of linear algebra to understand what is being covered in this course. If you are new to this subject or you have done your maths long time back, there is nothing to worry.

You can take elementary book on linear algebra, look up the concepts of expressing numbers or arrays in suitable matrix forms, understand about the dimensions of the matrices, how many rows are there, how many columns are there and what is meant by inverse of a matrix. You do not have to go into the detailed techniques of finding the inverse of the matrix. There are different tools available, different software available to find the inverse of matrices.

We have to understand what is meant by inverse of a matrix and you should also know about matrix addition, matrix multiplication, multiplying a square matrix with a column matrix. When you can multiply a square matrix with the column matrix, what should be the

dimensions of these square matrix and column matrices so that the multiplication is possible. So these are the basic concepts you should become familiar with.

I would imagine that it would take you maximum a couple of days to become familiar or refresh these concepts. If you can do that then the matrix manipulations will become very straight forward and we can carry out the regression analysis in a more efficient manner. So coming back to the multiple regression model, there are more than one independent variables. There are independent variables like  $X_1$ ,  $X_2$ .

These are called as regressor variables and  $\beta_1$ ,  $\beta_2$  are called as regression coefficients.  $\beta_0$  is the intercept and  $\beta_1$  is called as the partial regression coefficient 1,  $\beta_2$  is partial regression coefficient 2. The terminology is very important in statistics and statistical analysis.

**(Refer Slide Time: 09:38)**

**Partial Regression Coefficients**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$\beta_1$  and  $\beta_2$ : **partial regression coefficients** 1 and 2

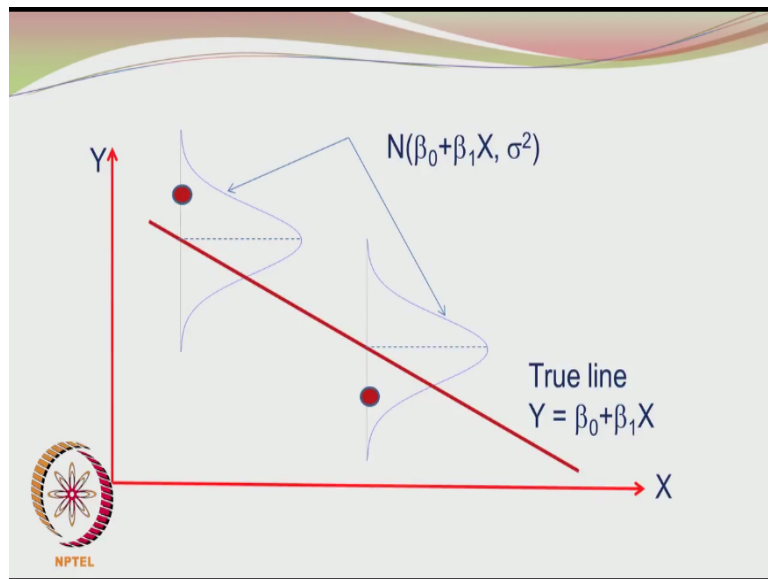
The term **partial** is used because

$\beta_1$  refers to the expected change in the response  $Y$  due to a change in  $X_1$  with  $X_2$  being kept constant.

NPTEL

So it is important to define them upfront.  $\beta_1$  and  $\beta_2$  are partial regression coefficients 1 and 2. The term partial is used because  $\beta_1$  refers to the expected change in the response due to a change in  $X_1$  with  $X_2$  being kept constant. What is the expected change in  $Y$  when  $X_2$  is kept constant and  $X_1$  is varied? Similarly,  $\beta_2$  is the expected change in the response due to change in  $X_2$  with  $X_1$  being kept constant.

**(Refer Slide Time: 10:16)**



This is a very famous diagram for linear regression. For simplicity, I have just shown 2 points. There are obviously more points but I am taking only a section of the diagram. If there are only 2 experimental points, I have shown beta 0 and beta 1 with only 2 experimental data or experimental observation so I could have fitted a straight line passing exactly through these 2 data points.

However, imagine that there are more data points and only a few of them are shown and my objective in linear regression is to make a line pass through such experimental data points in a way that it satisfies certain mathematical criteria. What are those mathematical criteria, I will explain in more detail later but essentially speaking the concept of least squares applies here.

Let us see, this is the deviation between the experimental data point and the model prediction line. Similarly, this is the deviation between the model prediction line and the experimental data point. So we are trying to balance the deviation between the data and the prediction. So what we do is we find the deviation between the experimental data point and the model prediction.

So that deviation is squared. Now we square the deviations of all the experimental data points from the model prediction and that sum of square of the deviations is minimized to find the parameters beta 0 and beta 1. You might have come across this already in least squares method. We are just using the same concept. The only difference is we will be doing with matrix manipulations.

So that large amounts of data can be handled efficiently. Let us assume that this line which is given here is the true line. In other words, let us say that it accurately represents the relationship between Y and X but the experimental data points are deviating from this true line and that is because of random effects. We talk about the error term being normally distributed and with constant variance.

So the experimental data points are lying anywhere around the true line here. The experimental data point can also lie here. It can lie anywhere as given by the spread around the true line. So the mean of this distribution is  $\beta_0 + \beta_1 X$ . This is also the expected value of the response. Now the data is scattered around it because of random effects and the variance of this probability distribution is given by  $\sigma^2$ .


Now when you come to the next value of X okay, this would be  $X_A$  and this would be  $X_B$ . This is the expected value of Y and this is where the actual data is lying that is because of random fluctuations. Again you describe a normal distribution around this mean value and you expect the actual experimental data to lie somewhere here. So what this really shows is the probability of the experimental data point lying further and further away from the true line becomes smaller okay.

We expect the experimental data points to lie closer to this straight line. This is the basic concept. It is important that these distributions describing the scatter of the experimental data around the true line is normal and these distributions have constant variance. Coming again the distribution describing the scatter of the experimental data from the mean or true line is normal.

And the distribution has a variance  $\sigma^2$  and all the experimental data points are also described by the normal distribution with mean  $\beta_0 + \beta_1 X$  and variance  $\sigma^2$ . All these distributions have the same variance.

**(Refer Slide Time: 12:53)**

- ❖ Each response is assumed to belong to a normal distribution centered vertically at the coordinate given by the regression line.
- ❖ The variance of all the normal distribution is assumed to be the same.



So each response is assumed to belong to a normal distribution centered vertically at the coordinate given by the regression line. The variance of all the normal distributions are assumed to be identical. The variances of all the normal distributions are assumed to be the same.


**(Refer Slide Time: 17:15)**

### Multiple Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

This is a multiple regression model with **k regressor variables** ( $X_1, X_2, \dots, X_k$ ).

The parameters  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  are called as partial regression coefficients.



Now let us be a bit more ambitious instead of talking about one independent variable or two independent variables. Let us talk about many independent variables or many regressor variables. So we have a mathematical model explaining the relationship between the experimental response and the independent variables. The unaccounted portion is due to experimental error or random error.

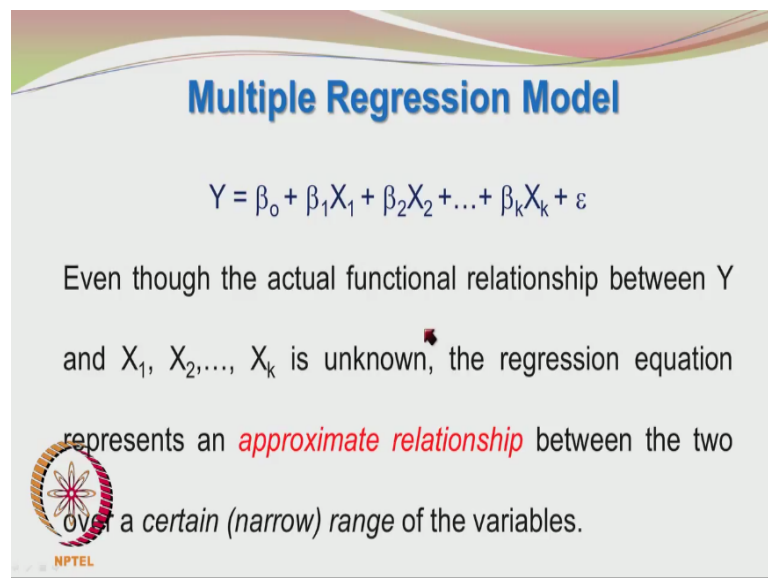


We hope that this particular model whatever we are proposing is adequate to describe the systematic dependency of Y with X1, X2, so on to Xk. So we have k regressor variables. Beta 0 is not associated with any regressor variable. The regressor variables are X1, X2 so on to Xk. Each regressor variable is associated with the coefficient such as beta 1, beta 2 so on to beta k.

Beta 0 is not associated with any regressor variables only beta 1, beta 2 so on to beta k are associated with the k regressor variables and these parameters are called as partial regression coefficients. What is a significance of beta 0? Beta 0 refers to the intercept okay. So let us imagine that you have a model  $Y = \beta_0 + \beta_1 X_1 + \epsilon$  and the model predicted would be  $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$  and beta 0 in that case would represent the intercept.

What is the response predicted when X1 goes to 0? Similarly, in a multi-regressor variable sense, when X1, X2 so on to Xk go to 0, beta 0 would then be the predicted value of Y.

**(Refer Slide Time: 19:50)**



**Multiple Regression Model**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

Even though the actual functional relationship between Y and  $X_1, X_2, \dots, X_k$  is unknown, the regression equation represents an *approximate relationship* between the two over a certain (narrow) range of the variables.

NPTEL

Please note that this is not a law, this is only an empirical model trying to or attempting to explain the dependence of Y on the different variables X1, X2 so on to Xk. For real life experiments, we may not know the actual functional relationship between the response and the influential factors.

Sometimes the process maybe very complicated and the equations describing the process may not be solved to give an analytical solution. In such cases rather than having a very difficult mathematical model, we try to understand the process behavior through a simple empirical

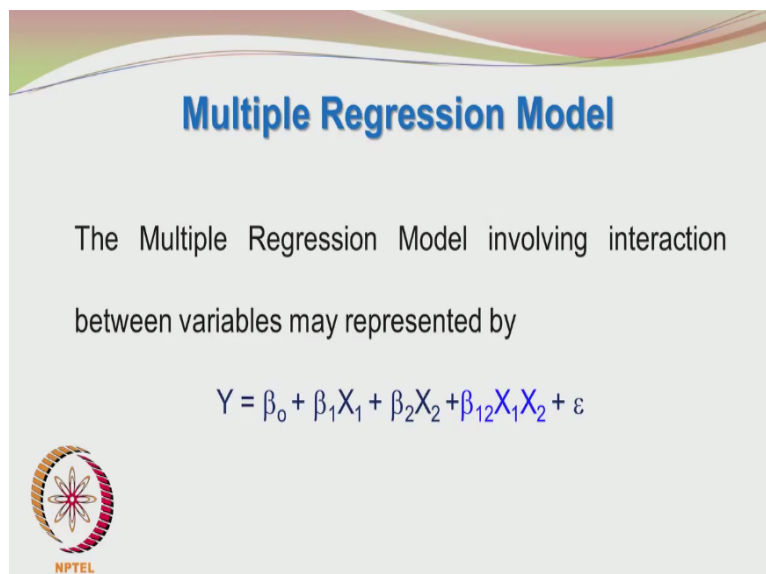
model. So the regression equation represents an approximate relationship between the response and the experimental factors over a narrow range.

When we do experiments, we do the runs only over a certain range. That range may be defined based on the limitations or okay that range may be based on certain constraints. You may not be able to achieve a relative humidity greater than 100 and you may not have a relative humidity less than 20% when you are doing the experiments. The temperature ranges in which you carry out the experiments may also be between 20 degrees to 70 degrees centigrade.

So these are defined ranges for your experimental variables and when you use the experimental observations, to develop an empirical correlation then please note that you cannot extrapolate the correlation to higher values or lower values than what you considered. This is very important because when you change the range of your experimental observations, the model parameters may also change.


You are assuming a certain relationship and that relationship may change. In fact, when you are trying to do calibration of instruments, you may find that the same calibration line may not apply over the entire concentration range. When the concentration crosses a certain value you may have to come up with the different calibration line. So the important thing to note is what is the range of the variables being considered in the present phase of experimental work and develop the correlation to account for the variations within this range.

**(Refer Slide Time: 23:33)**



**Multiple Regression Model**

The Multiple Regression Model involving interaction between variables may be represented by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$


NPTEL

The multiple regression model involving interaction between variables may also be represented by  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon$ . In the design of experiments, we saw that the interaction effects play a very major role sometimes even dominating over the main factors and here even interactions can be accounted for in the regression model.

We simply put a regression coefficient  $\beta_{12}$  and then take the  $X_1$ ,  $X_2$  variations into consideration and if you choose to put  $\beta_{11} X_1^2 + \beta_{22} X_2^2$ . So the choice of the model is yours okay. You can keep extending the model up to a certain point. You cannot extend the model indefinitely. The simple reason for that is you have certain number of finite set of observations.

And for solving any set of equations, you have to make sure that the number of variables is < the number of experimental observations. When the number of experimental variables is = number of experimental observations then you can get a perfect fit but usually when we have  $n$  experimental observations, the number of parameters we estimate from the regression model will be less.

So I will just explain this portion once again. As I was telling you, you can keep on adding more and more terms to this model but you cannot do so beyond a certain point. The important reason for that is you have to look at the number of experimental observations. If the number of model parameters is > the number of experimental observations, then you cannot find them.

If the number of model parameters is = to the number of experimental observations, then you will find an exact fit. Usually, the number of experimental data points is quite high, let us say 40 or 50 and the number of parameters you are estimating in the model may be 5 or 6. So the number of model parameters you are estimating should be smaller than the number of experimental observations.


It can also not exceed the number of experimental observations. So within these constraints, we can try the effect of adding more terms to the regression model.

**(Refer Slide Time: 27:00)**

## Multiple Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$

This is still a linear regression model as the functional relationship is linear in the unknown parameters ( $\beta_0, \beta_1, \beta_2, \beta_{12}$ ).



Just because you added  $X_1 X_2$  or  $X_1$  squared or  $X_1$  squared  $X_2$ , it does not make the regression model nonlinear. As I said before, the regression coefficients are still linear in nature. Here please note that  $X_1, X_2$  are not unknowns. The unknowns in this equation are  $\beta_0, \beta_1, \beta_2, \beta_{12}$ . So these are the unknown terms and they are all linear in nature okay.


So this one  $X_1, X_2, X_1 X_2$  can go even up to higher orders but as long as we have simple  $\beta_0, \beta_1, \beta_2, \beta_{12}$ , the estimation is still termed as linear regression procedure.

**(Refer Slide Time: 28:00)**

## Multiple Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$

Expressing  $X_1 X_2$  as equal to  $X_3$  we may write the above equation as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$


If you are confused by the notation  $\beta_{12}$ , you can simply define  $X_3$  as  $X_1 X_2$  and call  $\beta_{12}$  as  $\beta_3$  so that is what I have done here.  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$


error term. The error term as I said earlier may only be random error or it may also be the unaccounted part of the responses.

**(Refer Slide Time: 28:45)**

**Multiple Regression Model**

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{12}X_1X_2 + \varepsilon$$

Expressing  $X_1X_2$  as equal to  $X_3$  we may write the above equation as

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon$$


When you give the model to the software, we can get a 3-dimensional graph for this particular case once the regression model has estimated, we can get a response surface and that response surface need not be planar especially if this  $X_1$ ,  $X_2$  term is significant, the response surface may be a curve. It may even have peak if you have terms like  $X_1$  squared or  $X_2$  squared, there may be a maxima but it does not mean that the linear regression concepts are being violated.

The model parameters are still linear and the estimation procedure is called as the linear regression technique.

**(Refer Slide Time: 29:51)**

## Multiple Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Any regression model that is linear in parameters ( $\beta$ s) is a linear regression model, irrespective of the resulting response surface being a plane or curved.



So this is a more complicated model. Here the quadratic terms are being added in addition to the interaction term. You simply call them as  $X_3$ ,  $X_4$ ,  $X_5$  and use the same procedure to find the parameters or estimate the parameters. What is that procedure? I will come to it in a moment.

**(Refer Slide Time: 30:21)**

## Matrix Approach to Multiple Regression

Let there be  $k$  regressor variables with  $n$  observations

$$(X_{i1}, X_{i2}, X_{i3}, \dots, X_{ik}, Y_i), i=1, 2, \dots, n$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i=1, 2, \dots, n$$

Note:  $n > k$



This may be represented in a matrix notation as  $Y = X\beta + \varepsilon$

Now we are going into the matrix approach. A matrix is a 2-dimensional representation of numbers. Rather data are presented in an array form and this array comprises of both rows and columns. So what you need to do is in this array you have to define 2 indices  $i, j$ ,  $i$  refers to the row index and  $j$  refers to the column index. So if I say  $X_{ij}$  I am talking about a number which is present in the  $i$ th row and the  $j$ th column of the matrix.

Also you can have  $X_{ijk}$  then it becomes a 3-dimensional matrix but we are not going to look at such matrices in our analysis. We will be only looking at 2 indices. If you have  $X_{23}$  for example, it refers to the number which is present in the second row and third column of the matrix. Please note that  $X_{23}$  need not always be equal to  $X_{32}$  only in certain special matrices  $X_{23}$  may be equal to  $X_{32}$  otherwise the numbers may be unique.

Anyway that is the brief background on matrices. I am sure you will find lot more information in standard text books as I said earlier please do not go too deep into the subject, you just learn what is required for our present analysis. So we are having  $k$  regressor variables. What are the  $k$  regressor variables?  $X_1, X_2, \dots, X_k$  and you have  $n$  observations.

So you can represent this as  $X_{i1}, X_{i2}, X_{i3}$  so on to  $X_{ik}$  and  $Y_i$ . This may look a bit confusing. Let us look at only  $Y_i$  first.  $Y_i$  with  $i$  running from 1 to  $n$  represents the  $n$  observations of the experiment and  $X_{i1}, X_{i2}, X_{i3}$  so on to  $X_{ik}$  are required because for each experimental setting, you need one equation. So  $X_{i1}$  represents factor 1 or variable 1 for the  $i$ th run.

If you have  $X_{31}$  then it means the value taken by the first independent variable for the third experimental run,  $X_{31}$  is the value taken by the first independent variable for the third experimental run..  $X_{i2}$  or  $X_{32}$  in our example is the value taken second independent variable for the third run. So we have  $X_{ij}$  written with  $i$  running from 1 to  $n$  and  $j$  representing the  $k$  regressor variables.

So the model for the  $i$ th run or the  $i$ th experiment may be written as  $\beta_0$ . This is only an intercept. So we do not have any additional subscript here. This is universal to all the experimental runs.  $\beta_1, \beta_2, \beta_k$  are also universal to all the runs. We are not estimating  $\beta_1, \beta_2$  so on to  $\beta_k$  the regression parameters for each and every experimental run.

We are having a group of experimental data for the entire group, we are finding out the parameters  $\beta_1, \beta_2$  so on to  $\beta_k$  but the experimental conditions will vary for  $n$  experiments, you may have  $n$  different combinations of experimental conditions and that is given by  $X_{i1}, X_{i2}, \dots, X_{ik}$ . You have only independent variables or the factors influencing the experiment running from 1, 2 so on to  $k$ .

But these independent variables may take different values for different experimental settings and that experimental settings is given by the index  $i$  and that will run from 1, 2 so on to  $n$  and also please note that the number of experimental observations is usually  $>k$  the number of regressor variables. So now we can represent it in a matrix notation as  $Y=X\beta+\text{the error term}$ .

**(Refer Slide Time: 37:08)**

**Matrix Form of the Regression Equations**

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = X \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix}$

So you have  $Y$  which is  $Y_1, Y_2$ , so on to  $Y_n$ . This is a column vector, it is called a vector because it has only one column and it is having  $n$  entities so you can call it as a  $n$ -dimensional vector and so you have 1, 2, so on to  $n$ . These represent the different observations from your experiment. You have done  $n$  such experiments. Then you also have  $X$  which is the main matrix here.

This is not a square matrix, the first column in the matrix is always 1. Why do you need 1? In order to account for multiplication with  $\beta_0$ .  $\beta_0$  you please remember is not associated with any regressor variable. It is the constant term in the model equation without any regressor variable attached to it. It is the intercept for you to interpret it physically and so we have 1 here.

And then you have  $X_{11}, X_{12}, X_{13}$  so on to  $X_{1k}$ . What is  $X_{11}$ ?  $X_{11}$  is the value taken by the first regressor variable or the first independent variable for the first experiment.  $X_{12}$  is the value taken by the second independent variable for the first experiment.  $X_{13}$  is the value taken by the third independent variable or the third regressor variable for the first experiment,



so on to  $X_{1k}$  is the value taken by the  $k$ th independent variable or the  $k$ th regressor variable for the first experimental condition.

Now it is not necessary that all the values here should be different. In some cases, it may so happen that 2 independent variables are kept constant and the other 2 variables are varied or changed. So that is fine but there are some precautions you have to take that I will tell a bit later but what I am trying to say here is for a given experimental condition it is not absolutely essential that all these values taken by the different regressor variables or the independent variables should be different.

So how many such rows you will have? The number of rows you will have will correspond to the number of experiments run. So the last row will be  $X_{nj}$  where  $j$  runs from 1 to  $k$ . Now  $\beta$  is again a column vector and it is running from  $\beta_0$ ,  $\beta_1$  so on to  $\beta_k$ . You may think look the earlier defined  $Y$  the response vector, which was having  $n$  entities  $n$  dimensional vector but here is it not expected that  $k$  it should be  $n$  here not  $k$ .

We know that  $k$  is the number of regressor variables in addition to  $\beta_0$  but in order to make the matrix representation consistent, you should not  $k \geq n$ . The simple answer is it is not strictly necessary you can have  $k < n$ . How the matrices align themselves such that there is no inconsistency. We will see very shortly and then you also have the error terms  $\epsilon_1$ ,  $\epsilon_2$  so on to  $\epsilon_n$ .

So far the error terms is doggedly or persistently accompanying the regression equation.

**(Refer Slide Time: 42:07)**

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & \dots & X_{2k} \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & X_{n1} & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix}$$

$Y$  is an  $(n \times 1)$  vector of the observations  
 $X$  is an  $(n \times p)$  matrix of the levels of the independent variables

Now  $Y$  is a  $n/1$  vector of the observations,  $n$  is the  $n$  rows, normally when you are representing matrix dimensions, you give the row number first and then 1 is the column index. So you have  $Y_1, Y_2$ , so on to  $Y_n$  and  $X$  I explained all the terms in the previous slide. What is the dimension of  $X$  matrix? You have  $n$  rows and then you have  $k+1$  columns. So you have  $k$  regressor variables.

So you have  $k$  columns here and then  $+1$ , so we call  $p=k+1$ . So  $k+1$  is  $p$ , so  $X$  is a matrix with  $n$  rows and  $p$  columns where  $p$  is  $k+1$ .  $X$  is a  $n/p$  matrix of the levels of the independent variables. Beta is a  $p$  cross 1 vector of the regression coefficients and epsilon is a  $n/1$  vector of the random errors.

**(Refer Slide Time: 43:38)**

### Matrix Approach to Multiple Regression

$$Y = X\beta + \epsilon$$

We wish to find estimates for the linear regression model coefficients (also called parameters) which are expressed as follows

$$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$$

Now you have  $Y = X\beta + \epsilon$ . We want to estimate the parameters  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  so on to  $\beta_k$ . So when you are talking about estimated parameters or predicted parameters to distinguish it from the true parameters given by the  $\beta$  column vector, we put a hat to it to show that this is the predicted value. This  $\beta$  corresponds to a column vector of  $\beta$ 's, which are the true values.

But we do not know the true value from experiments. We can only estimate the values of the parameters from experiments and to show that these are parameters estimated we put the hat symbol. So  $\beta_0$  hat,  $\beta_1$  hat so on to  $\beta_k$  hat.

**(Refer Slide Time: 44:40)**

**Matrix Approach to Multiple Regression**

$$Y = X\beta + \epsilon$$

Using these coefficients in the following model will help us to predict the response for various values of  $X_i$  ( $i=1,2,\dots,k$ )

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$$

NPTTEL

Once you have these parameters estimated then you put forth a prediction equation and that is given by  $\hat{Y}$ .  $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$ . So now the error term has vanished okay. We are unable to account for the error term by this equation. This equation only gives the systematic variation of  $\hat{Y}$  due to change in the controlled factors  $X_1$ ,  $X_2$ , so on to  $X_k$ .

It does not explain the unaccounted or random phenomena. So that is why you have also in the predictive equation you are putting it as a  $\hat{Y}$  and then you do not have the error term.

**(Refer Slide Time: 45:49)**

## Least Squares Estimators of $\beta$

The least squares estimator  $\hat{\beta}$  is the solution for  $\beta$  in the equations

$$\frac{\partial L}{\partial \beta} = 0$$

Where



$$L = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon$$

Now you have to find the least squares estimator but before we go to the least squares estimator for the system of equations, I just want to take another look at the system of equations. So please look at these equations here.

**(Refer Slide Time: 46:25)**

$$\begin{matrix} \boxed{Y} = \boxed{X} \beta + \varepsilon \\ n \times 1 & (n \times p) & (p \times 1) & (n \times 1) \\ n \times 1 & n \times 1 & n \times 1 & n \times 1 \end{matrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_R \end{bmatrix}$$

$k > n$  No  
 $k < n$  Yes  
 $k$  need not be  $= n$   
 $k = n - 1$  Yes

$Y = X\beta + \varepsilon$  this is  $n/1$  and  $X$  was  $n/p$ ,  $\beta$  was  $p/1$  and  $\varepsilon$  was  $n/1$ . When you do matrix manipulations, you get  $n/1$  and  $p$  and  $p$  cancels, you get  $n/1$  and you also have  $n/1$ . What I am trying to say here is, the dimensions of this column vector should be the dimensions of the resulting matrix or vector here. This is not a column vector, it is an array comprising of  $n$  rows and  $p$  columns.

$\beta$  is also a column vector of  $p$  rows and one column. So when I multiply  $n/p$  with  $p/1$ , the  $p$  cancels out and I get  $n/1$ . So the given equation is consistent and even though  $\beta$  which

was having terms like beta 0, beta 1 so on to beta k, k need not be=n. Can k be<n? Yes. Can k be=n? Yes. This is a bit dizzy, we can say k=n-1 yes because you are also having beta 0 and k>n definitely no.

**(Refer Slide Time: 48:37)**

n observations  
 $p = k + 1$  parameters  
 n equations = n unknowns (max.)

No  $y_{des}$   
 $x = n$   
 $-1_{des}$

$$Y_1 = \beta_0 + \beta_{11} X_1 + \beta_{12} X_2 + \dots + \beta_{1k} X_k + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_{21} X_1 + \beta_{22} X_2 + \dots + \beta_{2k} X_k + \epsilon_2$$

$$Y_n = \beta_0 + \beta_{n1} X_1 + \beta_{n2} X_2 + \dots + \beta_{nk} X_k + \epsilon_n$$

So this is what you have to keep in mind because you have n observations or n experimental runs, you have  $k+1=p$  parameters and you have n equations. The equations are  $Y_1 = \beta_0 + \beta_{11} X_1 + \beta_{12} X_2 + \dots + \beta_{1k} X_k + \epsilon_1$ . You have the second equation  $\beta_0 + \beta_{21} X_1 + \beta_{22} X_2 + \dots + \beta_{2k} X_k + \epsilon_2$  so on to you have the nth data  $\beta_0 + \beta_{n1} X_1 + \beta_{n2} X_2 + \beta_{nk} X_k + \epsilon_n$  so this represents the n equations.

So you know that when you have n equations you can solve the n equations with n unknowns maximum okay. So that is what you have here, n unknowns maximum you can solve okay. Okay we will continue.

**(Refer Slide Time: 51:54)**

## Least Squares Estimators of $\beta$

The least squares estimator  $\hat{\beta}$  is the solution for  $\beta$  in the equations

$$\frac{\partial L}{\partial \beta} = 0$$

Where



$$L = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon$$

Let us now come to the least squares estimators of beta. We will discuss this in the next class.