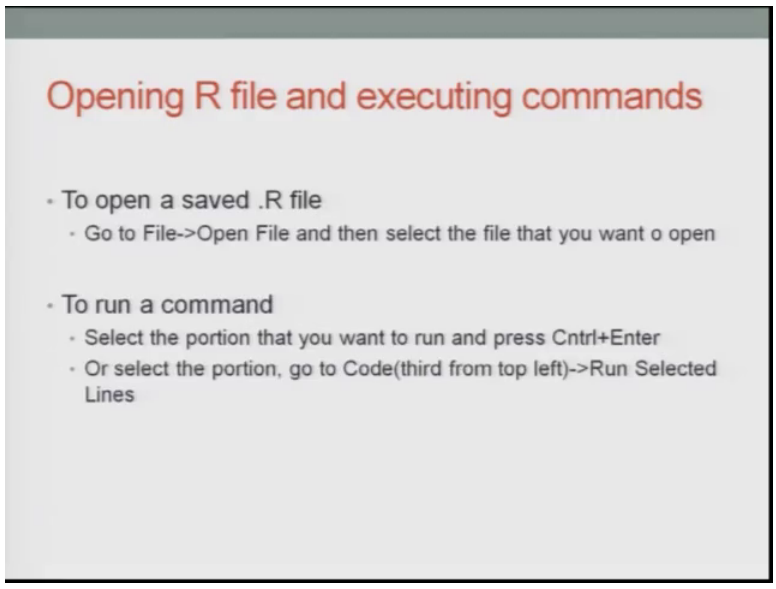
Total Quality Management - I
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 20
Basic Understanding of R and Introduction to Control Charts

Very good morning good afternoon good evening my dear friends; I am Raghunandan Sengupta from IME department IIT Kanpur. And this is the TQM 1 lecture and we are doing the course and we are doing the 20, 20th lecture today, which if you consider 5 lectures in one week. So, technically it should be the 4th week ending. So, as I was discussing in the all the statistical quality control charts and all these things can be very nicely done on in R. And as you know R is a free software you can download you can have from the R studios.

So, R studios if you look that it will give you a very good feel that it looks like a very interactive windows. Windows with all the commands can be recalled by the help menu, and it will really help you in basically trying to write small R commands and do the work and it can be taken up for on bigger scale also. So, to open the rs saved files wherever you save you basically go to the file in whichever directory it is been see it is whether you are saving in a desktop.

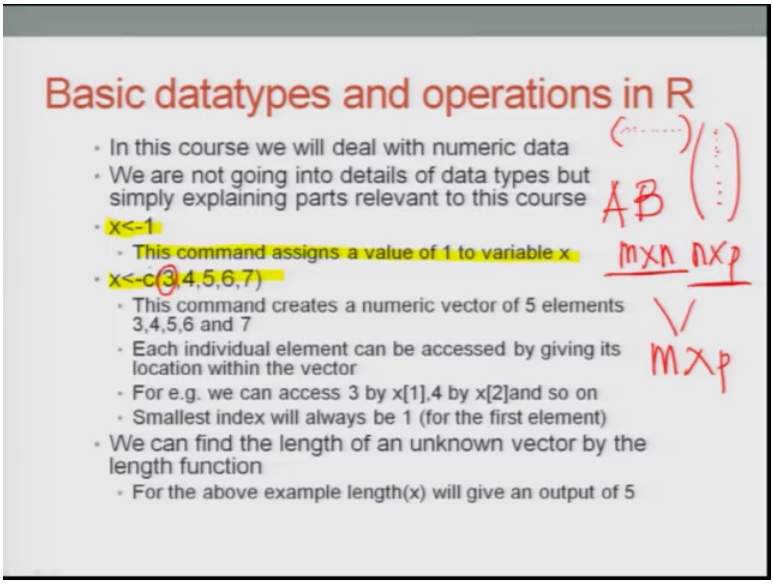
(Refer Slide Time: 01:18)



But best is that you see save under c such that all the programs for that for that laptop of the computer are all saved there. So, in case you want to retrieve delete whatever for the work which you are doing that should basically be safe saved in a separate partition which is d e whatever it is.

So, you go to the file open the file and then select the files you want to want to open or install. So, you run the commands select the portion that will run and then you basically startup the program control enter, and then the code can be run. So, once you run the code the results can be found out.

(Refer Slide Time: 02:04)



So, in this course we will deal with numeric data only that is TQM, and attributes even if they are there they can be converted into new numerical framework. We are not going to in details of the data type like whether they are stacked whether they are randomized all these things informations, we are not going to discuss. We will presuppose the information which is there has already that set of information already there.

But simply explaining parts relevant to the course would be basically. If you assigned with a command let me highlight it will be better if I. So, if you use the command this one. So, this command basically assigns a value of x to the variable one or whatever you want to if you basically may create a column vector or a or the row vector. The command is this and this command creates a numeric vectors of 5 elements whether you want to

would have in a transpose, whether you want to have a row that will depend on what type of a matrix multiplication which is there.

So obviously, if I if I would write try to write it in red color. So, say for example, they had 2 matrices A and B. So obviously, in the size of this matrix should be such that matrix multiplication should be easier. So, if you are saving n ma it should be of size m cross n. And B should be size of n cross p. So, as the overall multiplication gives you a matrix, which is whatever it is has a size of n cross c. If you are multiplying a vector with another vector, when if it is a dot product or a cross product; obviously, the multiplication should be done accordingly.

So, if it is a dot product obviously, it will give you a 6 scalar quantity. So, in this set the vectors all the elements in the first one will be along the rows as I am pointing and all the elements for the second vectors would be along the columns. So, once you multiply element one with the element one element 2 with element 2 and also so on and so forth. You will get one element which basically would be the scalar. So, in this command this where do you have a column this command creates a numeric vector of 5 elements which are 3, 4, 5 and 6 each individual elements can be recalled or accessed by the location in the vector.

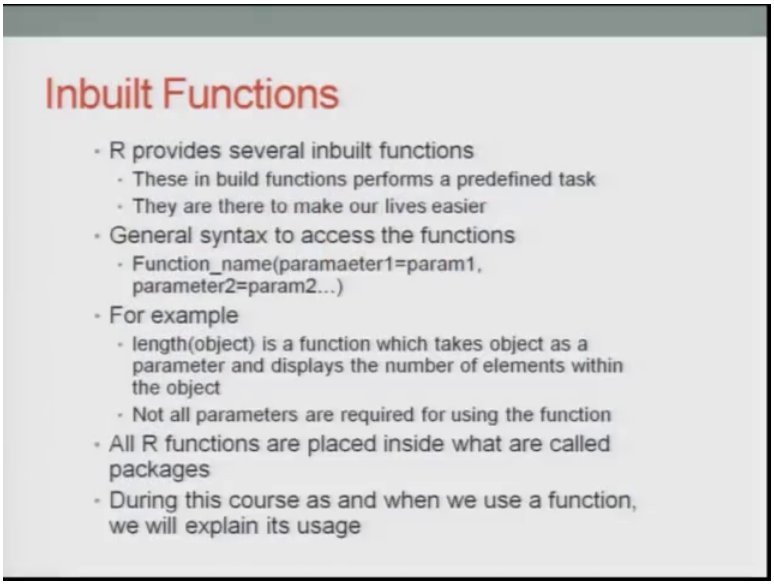
So, if you consider the number 3. So, basically it is 1 cross 1. So that means, there are only 5 of the actual vector if it is this like a row 1. It basically has 5 number of columns and one number of rows. So, if I am basically mentioning the rows. And the columns depending so obviously, it will be 1 cross 1 for number 3 it will be one cross 2 number 4 for number 4 and so on and so forth. So, this command creates a numeric vector of 5 elements each element can be accessed by given x location within this vector. For example, we can access 3 by giving the number x 1. 1 means the first element 2 can be given by 4 can be given by 2 5 can be given by 5 and so on and so forth.

So, if it is basically matrix; obviously, the cell position depending on the row and the column would be specified accordingly. The smallest index would always be when you are basically trying to analyze the index concept. So, the smallest index would be 1 which would be for the first element we can find the length of an unknown vector also by the length function and we can and if you want if you have a vector or us or a scale or a

row vector or a column vector accordingly. You can specify the length of x which is the vector and it will give you the number of elements which are there.

So, in generally if somebody is aware of matlab. So, that is basically given by the size. So, if you have if I specify the size of the vector. Then I am able to get what is the number of rows and what is the number of columns.

(Refer Slide Time: 06:17)

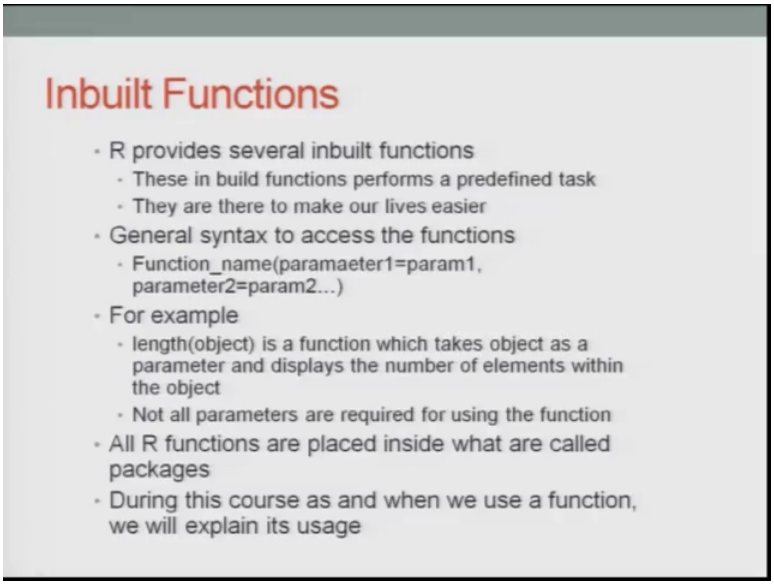


So, there are inbuilt functions. So, R provides several inbuilt functions these inbuilt functions perform a predefined task. And they can be called record from the library they are here to make our life easy. And basically we can use those inbuilt functions or library functions to do our calculations accordingly. So, general syntax can be basically we have to define a function with this corresponding parameters can technically that I want to basically find out the humidity and the and the temperature.

So, I specify the parameter. So, the parameter can be discrete parameter can be continuous parameter can basically be discrete between say for example, 0 and 10. So, I have to basically specify that accordingly if it is a continuous 1 also I have to basically do the needful in order to specify the same. For example, the length of an object is a function which takes an object as a parameter and displays the number of elements within that object. So, in that case when here is a vector you are trying to define it as an object which will give you the length means number of elements which are there.

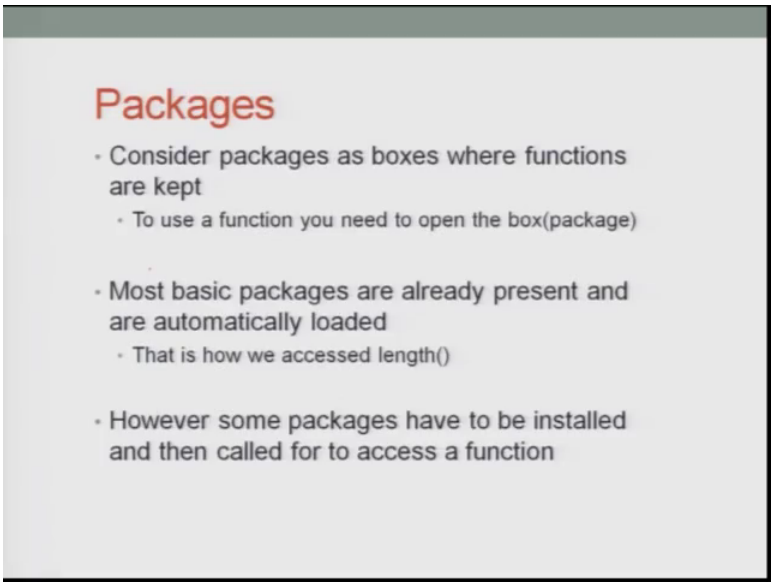
All our functions are placed inside what are call the packages. So, if you consider the R as a software. So, all that packages for different type of library functions would be combined under the so called sub directories. So, the word subdirectory is not true, but basically they are clustered and all are all on different domain areas So, you can access them accordingly.

(Refer Slide Time: 07:44)



During this course as and when we use a function we will explain it is uses accordingly.

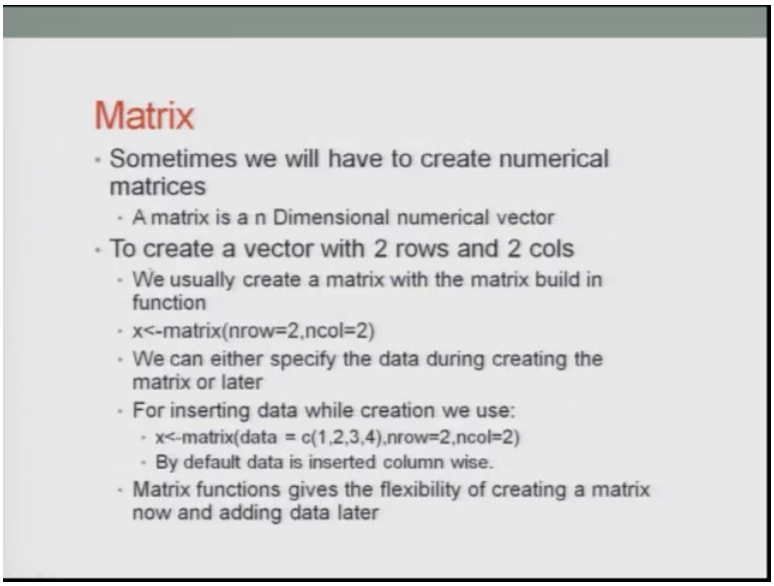
(Refer Slide Time: 07:49)



So, other packages R for the rs would be consider packages as boxes as I main subdirectories where functions are kept. To use a function you need to open that back box or a package or subdirectory and recall the functions accordingly.

Most packages are already present in R and they are being developed because R is an open software they are being developed accordingly. So, that is how we create access the length as the vector; however, some packages has to be installed and those inbuilt functions has to be recalled. So, if some packages are not there you have to basically load them from the net download them, install them in your actual R software which is already existing and then run them accordingly.

(Refer Slide Time: 08:32)



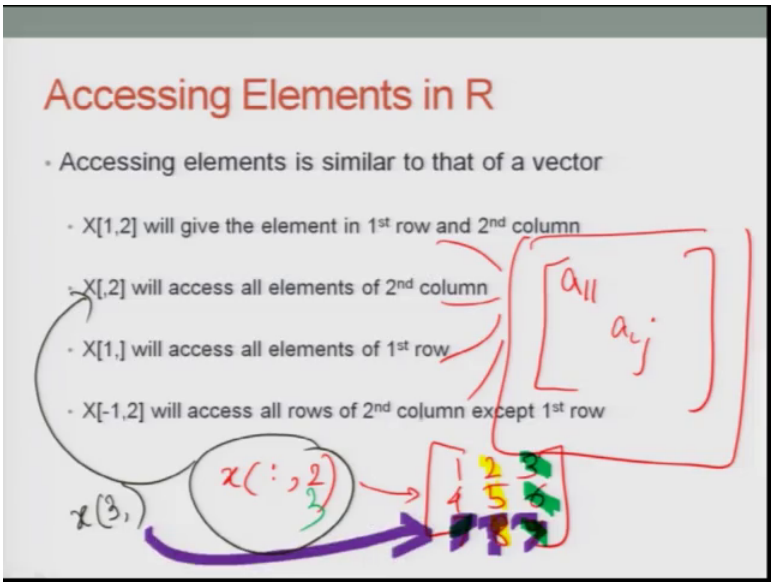
Coming back to the matrix notation which was discussing. Sometimes we have to create numerical matrices also a matrix is basically our dimension of vectors. So, it will be m cross n depending on the rows and the columns.

So, to create a vector with 2 rows and 2 columns you basically specify. A matrix that is the syntax when you are trying to basically may specify the matrix. So, and we name the number of rows and n rows is 2 that will create the number of rows as 2 and it will create a number of columns as 2. So, once it is a 2 cross 2; obviously, you can input the data and to fill up the matrix accordingly it can be done. So obviously, remember the first element would be the number of rows second element would be the number of columns.

So, you can either specify the data during creation on the matrix. Depending on how we want that it can be that is a is a matrix which has got empty spaces and we filled us fill up the cells with the numbers accordingly, or it can be predefined and we fill up the cells with numbers predefined. For inserting data you basically while creating we use the matrix and the datas are 1, 2, 3, 4. So, it means that they would be in certain and if you have already specified the number of rows and number of columns as 2 and 2, and you num mention the numbers 1, 2, 3, 4. So, they would be inserted column wise.

So, the numbers would be filled not row wise, but column wise as that the number sequences. Whatever you mention in that that set c would basically be the numbers technically would picked up from set c and put into the matrix column wise and then it would be done accordingly. Matrix functions give the flexibility of creating a matrix now and adding the part of the matrix later on or part of the rows later on such that the a matrix can be expanded or contracted; that means, size can be made larger. So, if it is m cross n you can make it say for example, m plus 3 into n plus 13. So, then extra 3 number of rows and extra 13 number of columns can be added accordingly to make the calculations as per the information which is given.

(Refer Slide Time: 10:46)



So, you want to access the element which is in the first row and the second column as I am mentioned you have to basically mention the cell. So, if I have a matrix. So, what I am trying to do you notice this a one is the first row first column. If I am mentioning aij

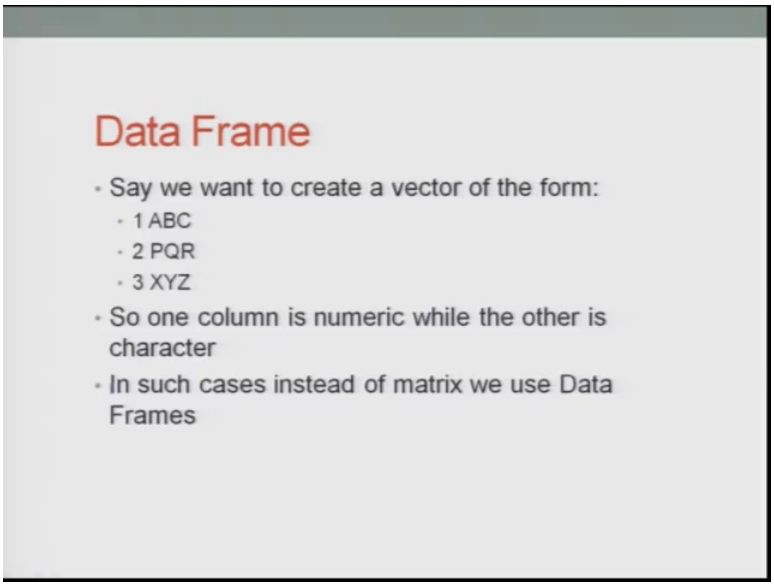
it is basically means the I th row and the j th column. So, whatever I mentioned here by my writing would immediately make sense for the 4 bullet points which are mentioned. So, x 1 comma 2 gives you the element in the first row and the second column x sum colon. So, here in matlab generally we will be using x colon comma 2.

So, this means it will take access all the elements which are in the second column. So, say for example, my matrices are 1, 2, 3, 4, 5, 6, 7, 8, 9. So, if I mentioned a x for the matlab code and I am going to come to the R code within a few seconds. So, this would mean that it will access let me highlight it would be easier for us to consider. It will access all the elements which are in second column. So, which are as you can see are the values 2, 5 and 8. If somebody says in place of 2 it is the number say for example, 3 here. So, in that case the values would be. So, if it is x colon comma 3. So, the numbers would be 3, 6 and 9.

Now, if I go into so obviously, the information which is given here. This one would immediately make sense if you go here which means x blank comma 2 would access all the elements in the second column. If I make the information x j comma blank it will access all the elements in the j th row. So, if I make a for information in this set. So, it either this, which will now access that we change the color if. So, this would be 7 8 and 9. So, this will go as for them.

Now, x 1 comma 2 will access all rows of the second column except the first one. So, one I mentioning minus 1 it will eliminate that. If I am in mention minus 1 to minus 2 it will eliminate the first one first one and the second one rows, and I can access all the elements which are in the second column starting the third 4th fifth 6th row whatever it is. And it can be made accordingly for the accessing limited number of columns while we are accessing all the rows at one go.

(Refer Slide Time: 14:10)

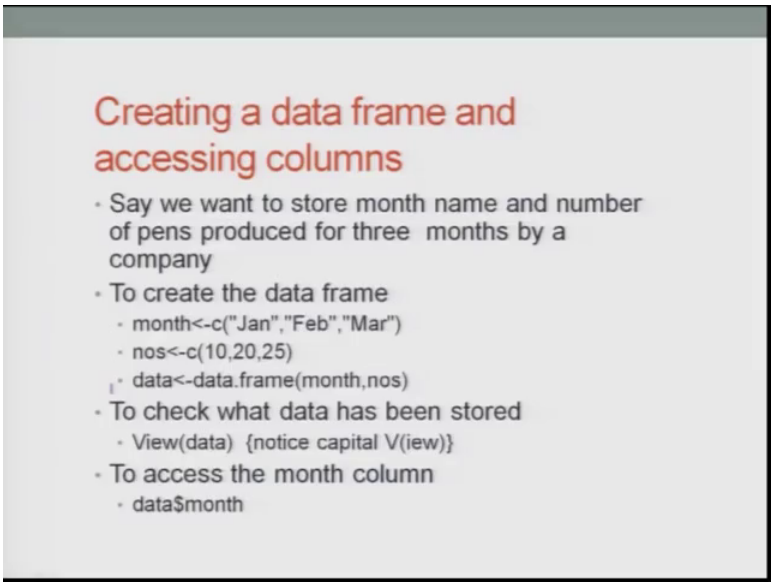


So, when you are talking on the data frame say we want to create a vector of the form of like abc or pq or xyz. So, one column is numeric while the other one is character So that kind is possible.

So, in such cases instead of the matrix we use the data frameworks. So, the first set or sets of rows or first few sets of the columns. So, the columns and rows can be at any place; that means, those which are numeric. So, it can be second third 4th or it can be say for example, fifth 7th ninth for the column for the rows, or the columns it can be made accordingly and we can make the matrix as a at a data frame and basically put input the data which is a combination of numeric and alphabet as may be required for the study.

So, if you remember in the house of qualities you mean mentioning that rather than giving some points we may assign some numeric, along with alpha numeric values depending on what is the overall feedback which you are getting between the combinations. So, we can basically use those the data frames in order to access the data and put input the data.

(Refer Slide Time: 15:24)



So, creating a data frame and accessing columns would be say we want to store month name. So, January February march April month name and the number of pens being produce for this the months accordingly. So, to create the month a data frame we defined in c as January February march. So, these are the c months and for the numbers corresponding January February march through the total number production what we mentioned has 10, 20, 25 whatever it is.

If it is more than one product like pen of red ink pen of blue ink pen of several green ink the blue ink and accordingly. So, you can basically have for January first column then for the corresponding values of February and march in and April and so on and so forth, can be input in say for example, the second column the third column and 4th column. So, what you will actually have is this. So, you will have a matrix on the data frame. So, the months are written here. And the values of pen of p one which is red pen where is a p 2 which is blue pen accordingly will be put up in the second column the third column and the 4th column.

So, this whole thing is the data frame which will you place accordingly for our calculation to wake. To check whether the data has been stored you can basically view the data and that the concept if you remember that this whether it is a small letter word or a capital letter word would make sense because they are case sensitive. So, the word view in matlab is just for the information and you can check up those in the

corresponding library functions also view is basically with capital v and you can access the data and have a look at the matrix of the data frame. To access the month wise column you will basically input the data and at a fix month.

So, if you remember in excel you use the dollar sign to fix that cell or fix that column you will use that accordingly to access the data from R. So, basic few basic mathematical operations and sequences are. So, if you let x be given a value of 4 y being value of 3. So, when you find out the sum. So, it will be given by 7 if you find the difference it is given by 1, multiplication is the star 1 which will be given by 12 if you want to find out the ratios of x to y; y to x you will get the values accordingly. So, you want to generate numbers from one to 5. So, you can write them as the column as one to 5 and the shortcut would be depending what is the range or the domain space.

So, 1 colon 5 or 1 colon 10 would generate the numbers depending on how many such numbers you want. So, you can also use the sequence function. So, sequence function would give you that what is the start number what is the end number and what are the step function. So, that they are very has a helpful in trying to generate the numbers for the discrete distributions and so on and so forth, even for histogram preparation and accordingly. In matlab in R in all functions you obviously, you have different type of loops if and when loops do while loops so on and so forth.

So, they basically they do the calculation time and again. So, for the loop function a loop is used when you want to repeat some sequence of step or sequence of functions in a or for a finite amount of time. So, depending on the counter of I changing from 1 to n or j changing from 1 to k you can basically have a different type of loops accordingly, but only remember which is applicable for all programming and all logic is that the loops should basically have a sequence as a logical concept or basically coming back to the source or from where they start, which everybody is aware, but still I would like to mention because that becomes very a very problematic when the loop is not working.

And then you find out the logic of the sequence of the loop is syntax wise is making sense. But when you go into the logical framework it does not make pixels like say for example, if you have a do while loop then; obviously, any other loops of if then loops should come in such a framework that the concentration of the loops are done such that one loop is operated, then only the final results of that inner loop is basically given as an

as an input for the second layer and this output of the second layer is given as an input for the third layer and so on and so forth.

So, there is no cutting of the loops such that if there is a committee cutting of the loops or (Refer Time: 20:17) crossing of the loops then it becomes a problem that the logical sequence is broken. So, this is a very important point which I want to mention not specific to R not specific to matlab, but these problems occurs when you are basically trying to make programming for the study of whether TQM 1 or TQM 2 whatever it is. So, be very careful about that. So, when you are making the syntax, the syntax would be for conditions whatever the for conditions are conditions can be repeated and the sequences of the steps can be repeated accordingly depending on the or the if and then do while loop for conditions.

So, let there be example consider that x is 3, 4, 5 and you want to add one to all of them. So, it 3 would become 4, 4 would become 5, 5 would then 6; 6. So, what you do is that for recall the number in that is in that vector, vector means either row vector or a column vector and basically that would be repeated 3 number of times. So, that is why it is given as a length of x, and then that x I th one. So, x is the is what is the number which is being called and which place is basically denoted by I. Now I would take all the values starting from 1 till the length. So, when I is 1 it will be 3 3 would be added with 1 it becomes 4 when I is 2 1 the once the first a step is completed that I increases by 1 unit and becomes the second one.

So now you will the computer will ask as per the logic that what is x 2. So, x 2 is now 4 4 been incremented by one and becomes 5 one this whole sequence is over then I would be in increased to the next level which will be 2 plus 1 3. For as it is doing an each step it is basically trying to find out whether it is exceeded or it is at the limit or what is the length of the column. So now, it sees that the column length has been reached. So, which is the third element. So now, I will be 3. So, corresponding the number x 3 would be called which is 5 and x 5 would be incremented by 1. So, hence it will become 6.

So, we will come back to the packages R later on as we do the problem. So now we will start with this basic statistical process control tables and charts. The statistical process control or which is spc basically monitors the production of the defect and tries to prevent poor quality. We are interested to find out some sample from the total population

and use control charts to study how the sample is performing with respect to the production process. And try to find out whether than aberrations which needs to be corrected.

Now, variabilities are is basically inherent in any process they can be natural or common cause, special assignable causes would also be there, but we need to take corrective actions accordingly. So, they provide a statistical signal when assignable causes are present and assignable causes can be studied in more details as thus corrective actions can be taken. We need to basically detect and eliminate the assignable causes which are the causes of variance so obviously, white noise would be there. Which is non assignable causes which are inherent in the system you may not be able to do away with that when you are basically taking a random reading; obviously, there would be variation.

So, what we want to understand and what is the essence of total quality management or quality control is basically to limit the overall assignable causes as low as possible and try to take preventive actions for the same. Now variability can be coming from random causes which can be common causes inherent in a process. They can be eliminated only thing improvement in the systems. And non random causes would be some special causes, they are due to identifiable factors which can be identified when correction steps should can be taken. Can we modified through operator or management actions hence the quality over overall scheme can be improved.

Random or natural variations are a natural variations in production process are there, these are to be expected. So, output measures can follow a property distribution. For any distribution there is a measure of central tendency and obviously, their dispersion, but you should be aware and take corrective actions accordingly. Non random or assignable variations are there. So, variations that can be traced to a specific reason like the flow of coolant is not proper or there are some a problem in the machinability of the machine or the humidity is very high, or temperature is very high, or the raw materials which is being used has some problems. So, those would be some defects would be coming out which we can take correct corrective actions accordingly.

Variations that can be traced to a specific reasons. Like as I mentioned machine wear misadjusted equipment fatigued or untrained workers and all these things. The objective is to discover those assignable causes and take corrective actions and eliminate them

accordingly. Now quality measures can both be variables and can be attributes attributes are some characteristics which technically you will give some qualitative definition and variables are quantitative one which you can study. So, attribute is a is a characteristic with the product that can be evaluated with the discrete response good, bad or hot, a very hot, cold, very cold, humidity, percent very high, average low and so on and so forth they can be studied. Well a variable is be all characteristics of product that can be continuous or that can be measured.

So, it can be weight length humidity it can be temperature it can be speed whatever we want to study. So, the types of datas would be as I said variables. So, they be characteristics that can be taken a real values, may be in whole or in fractional numbers they can be continuous random variables and they can be discrete also we will study that later. Well on the other hand for the attributes are their defect related characteristics. So, those are characteristic some attributes some qualitative information. They are able to classify products at either good or bad or on count the numbers. They can be I am trying to find out color of some scheme or and if the color scheme matches well; obviously, we will see the they are in category one if the color scheme does not match we will see in category 2.

So, those color schemes where you are not able to give some assignable variables or numbers to that. They are categorical or discrete random variables and they can be studied accordingly. So, can for charts for variables are for variable lengths we can have continuous dimensions, they can be weight can be speed can be length can be strength they can be say for example, I want to understand what is the fluidity what is the youngs modulus when we are basically trying to do some mechanical test it.

So, those can be studied. Technically we have 2 type of charts. One is the x bar charts which are to control the central tendency or the pros overall process. So, how variability is high how variability is low. And there would be R charts which are to control the dispersion of the overall process dispersion I am I am talking from the point of view of very variability or variance, but we will try to basically find out what is the range are what basically means that is based on which you can do this study.

So, if I if I need to understand how the charts look like, we need to for controlling the mean in the variability. Some of some in the charts can be say for example, in the one

which you I am circling you basically I am some mean value there is some upper control limit lower control meet which I am now marking. And they would be a mean and standard deviations would be measured at a nominal value and we study accordingly. For the other 2 charts which are given we are trying to basically find out the change in the mean process where the mean process is shifting that is mu naught is shifting. And another would be how fat or how thin the charts would be.

So, these charts or tables are or graphs are all normal distribution. I will come to that normal destruction later, on how they would give you the information based on which we can use the R and the x bar charts for the variability study in the in the process control. With this I will end this lecture and continue with the concept of studying the process control charts in the laters classes.

Thank you, and have a very nice day.