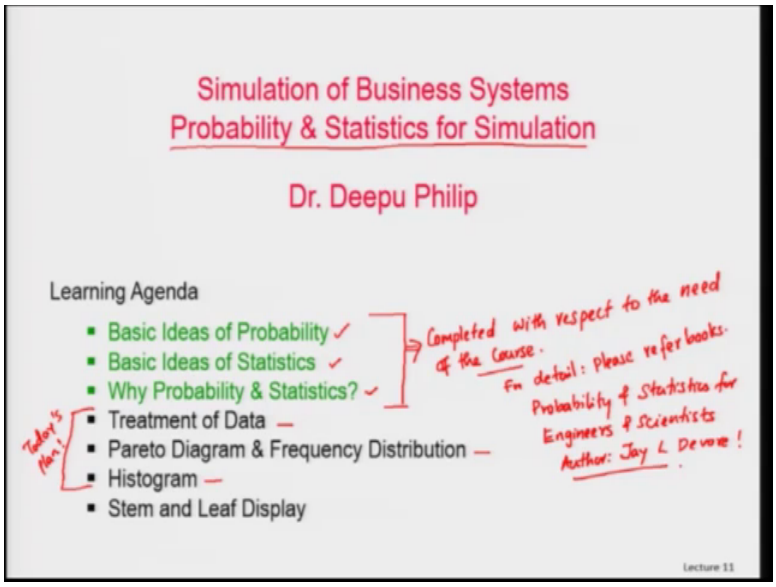
Simulation of Business Systems
Prof. Deepu Philip
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 14
Pareto Analysis

Good morning students, welcome to yet another lecture of Simulation of Business Systems.

(Refer Slide Time: 00:19)



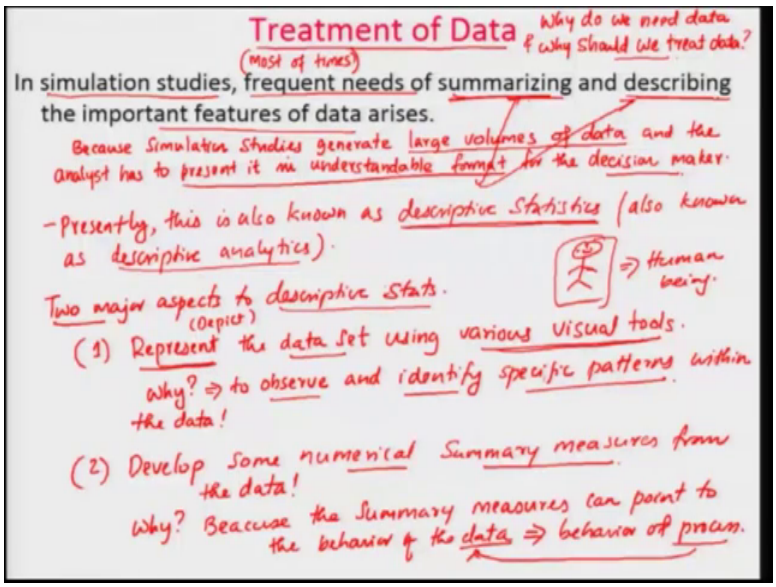
And we are in the MOOCs course on this advanced technique of an studying systems, business systems typically productive systems and today, we are going to cover the topic on continuation of the topic on Probability and Statistics which is necessary for simulation. We are not talking about classical probability and statistics only the portion which is specifically required for this course and I am Dr. Deepu Philip, I am from IIT, Kanpur.

So, if you look into today's agenda, the Probability and Statistics for Simulation was a big broader topic. We have already gone through basic ideas of probability, we also gone through basic ideas of statistics. These are and then, why probability and statistics for engineers and business analyst or business manager. These are as if you think about it, we have completed this with respect to the need of the course ok.

So, whatever the course requires for the time being is what we are covering. We are not really covering everything in detail. So, for detail, please refer books ok; Probability and Statistics for Engineers and Scientists by J L Devore is a pretty good book and he can; so, the name of the book is Probability and Statistics for Engineers and Scientist; author is J L Devore ok. This is a good book for people to study more of these things in detail. We are only looking at simple things that are pertaining to this particular topic of this course.

Today, we will try to cover these concepts of Treatment of Data; why it is important and why it is necessary for simulation. We talk about 2 simple tools Pareto Diagram and Frequency Distribution and then, we will also go to a Histogram. We will try to cover all these 3 in today's lecture ok. So, we will say this is our today's plan; we might finish it in 1 lecture or we might finish it in one and half lecture kind of a thing ok.

(Refer Slide Time: 02:59)



So, the first topic or the first part we need to do is we talk about this Treatment of Data ok. Why we do we need to the question here is why do we need data and why should we treated data ok. Towards the end of this lecture we hope that we would be able to understand or get an answer to these 2 particular questions. So, in simulation studies, specifically in simulation studies the frequent needs arise that too lot of needs lot of the time; so, you can say frequent needs means ok, most of times most of the times you require summarising and describing important features of data ok.

So, your aim in the simulation studies is to summarise ok. Summarise mean provide a problem summary and why do we need to do this? Because Simulation studies generate large volumes of data and the analyst has to present it in understandable format; format for the decision maker say like a CEO, CFO those kind of things.

So, the decision maker does not have time to go through these large volumes of data, huge amount of data. So, in the job of the analyst becomes summarising and describing the data ok; presented in the understandable format involves these two ok. So, summarise and describe the data. So, some of the things that we are going to study in this particular class is about how do we summarise and describe the data.

Presently, this is also known as this is also known as Descriptive Statistics ok. So, this process of summarising and describing important features of data is commonly known as descriptive statistics; also known as known by the fancy term descriptive analytics ok. So, this descriptive statistics and descriptive analytics, can we use the terms interchangeably ok. As far as I am concerned that is not much change in this except the fact that there is some people will talk about their there is a hypothesis, but for the practical purpose of this course we can understand, we can assume that the both are exactly the same ok.

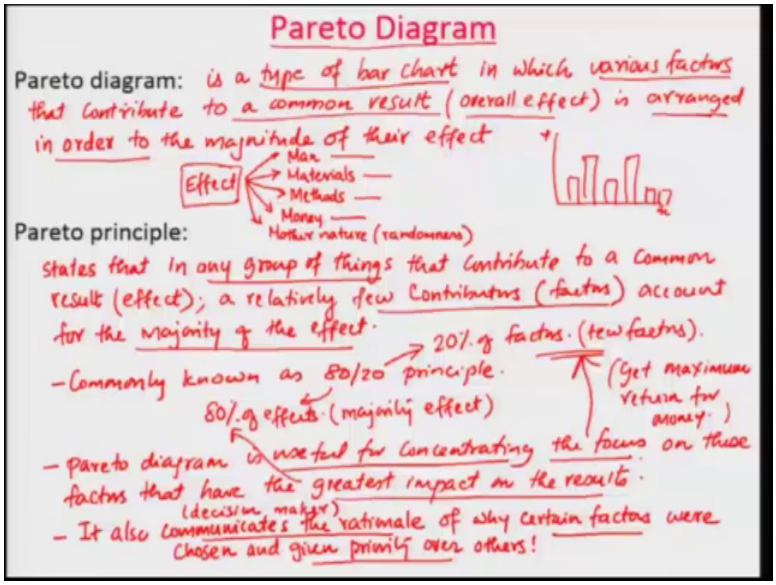
So, this type this field of statistics that deals with summarising and describing the important features of data is called as the descriptive statistics ok. We will say there are 2 major aspects to this. 2 major aspects to descriptive statistics ok. There are 2 major aspects. What are those 2? Number 1, first one represent the data set using various visual tools ok, so, the first aspect of the descriptive stats is you represent or you depict or you depict the data set using various visual tools ok. Why do we do this? The question is why do we do this? The answer is to observe and identify specific patterns within the data within the data ok.

So, what we do is when you use various visual tools because people are visual in nature we are very good at identify. So, if I draw a picture like this, you will say it is a human being ok. If you do this, you will say it's a smiling human being; whereas, it is not even remotely close to what are actual human beings looks like. So there is a stick figure, but we are able to identify various patterns, specific patterns, patterns that are with respect to a human being. So, we are able to identify and say this as a human being. So, same way,

a presentation of human being. So, that same way we are able to observe and identify various patterns within the data and for that visual tools are very useful.

The second aspect of is develop some numerical summary measures; numerical summary measures from the data ok. So, here what we try to do is we try to develop some numerical summary measures or try to summarise the data. Why? Because these measures, the summary measures can point to the behaviour of the data which implies behaviour of process; why? Because process generates data ok; data is generated typically by a process and if you can summarise the data, summarise numerical measures from the data; you are actually summarising the behaviour of the process. This is the reason why descriptive statistics is very important for the simulation studies.

(Refer Slide Time: 10:19)



So, one of the first tool that we will study in this class or this concept is the Pareto diagram ok. Pareto diagram is quite common quite famous and lot of people uses that and for multiple reasons and what is a Pareto diagram? It is a Pareto diagram; it is a type of bar chart ok, a bar chart in which various factors in which various factors that contribute that contribute to a common result, to a common result ok.

Some people also call this common result as overall effect is common result or overall effect is arranged is arranged in order to in order to the magnitude of their effect ok. The major points of this is it is a type of bar chart ok. Bar chart is something like you have x axis y axis and you have bars like this ok. We can call this as a bar chart ok. A type of a

bar chart in which various factors; so, we are looking at various factors pertaining to the data that we have there contribute to a common result or an overall effect. So, there is a common result to which various factors are associated with ok.

So, and a way to look into this is an effect or a common cause ok, effect can have multiple factors ok. So, for the time being I am going to give you 5 main ones; man, materials, methods, money and obviously, yes something called as mother nature or randomness ok, we call it as uncertainty. So, man is the human being there is involved; materials is the materials that are using to make the process; methods is the practices that you are following; money is the finances that are required ok. So, all those things are important.

So, these factors contribute to the a particular effect and you are trying to find out which one of these factors contribute to that particular effect the most, that is the idea here ok. So, there arranged in order ok; arranged in ascending or descending order or appropriate order; arranged in the order of magnitude of the result ok. So, what is Pareto principle? There is a this Pareto diagram is based on a principle called Pareto principle.

This states that states that in any group of things, in any group of things that contribute to a common result that contribute to a common result to a common result states that in any group of things that contribute to a common result or the other word is an effect ok, result is also people called the Scientific Emerson Effect, a relatively few contributors, few contributors or we can talk about the factors; contributors are the factors; obviously, contributors account for account for the majority of the effect.

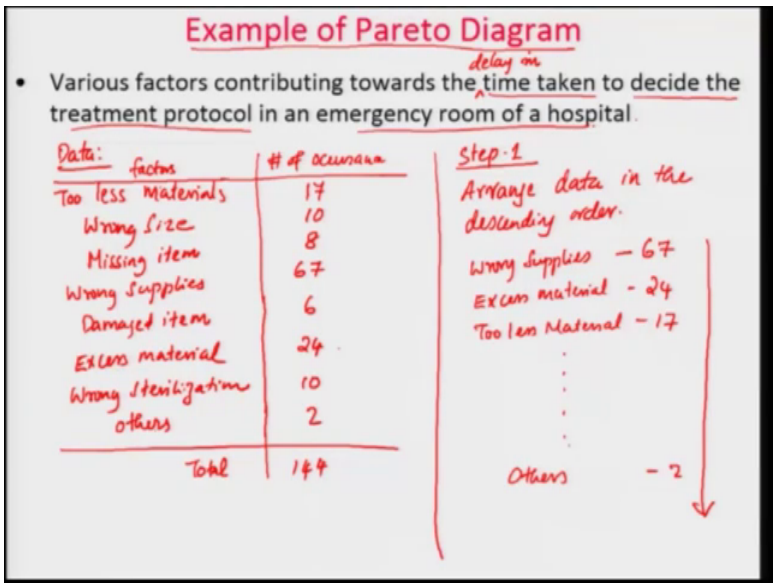
So, what we are saying here is that a small number, a relatively few small limited numbers of them account for majority of them or it is commonly known as a commonly known as 80-20 principle which means this is 20 percent of factors, contribute to 80 percent of effects ok. So, the 80 percent of the effect or the result is basically contributed by 20 percent of factors, this is the few factors and here is the majority effect ok. Hope it is clear to you guys now ok.

So, what does it do? So, then Pareto Diagram is useful or is useful for concentrating; it is useful for concentrating the focus on those factors that have the greatest impact on the result on the results. So, what we are saying here is it is useful for concentrating the focus.

So, we can pretty much focus on these limited few factors; focus on these 20 factors or twenty percent of the factors which have the greatest impact on the results or which has 80 percent of the impact on the results, focus on them. So, that we can get the max. So, the aim is to get maximum return for money ok. So, if we focus on the factors that are contributing to the most of the majority of the effects. Then, obviously, if you change them you make them better then your production will also be better ok.

Additionally, it also communicates, it also communicates the rational of why certain factors; why certain factors were chosen were chosen and given priority over others. So, when you are trying to talk to the your decision maker ok, the communication is with the decision maker ok. So, the decision makers when you are trying to tell them why you are focusing on certain factors and why are you giving priority of to those factors over others. For that, your Pareto diagram is a good communication tool.

(Refer Slide Time: 18:12)



With this what we will try to do is we will try to work a small example of Pareto diagram and see how this is being constructed and so, now, let us see how we can build a Pareto diagram ok. The various factors case the example problem is that there are various factors that contribute towards the time taken to decide the treatment protocol in an emergency room of a hospital. So, the factor here is time taken to decide the treatment protocol in an emergency room of a hospital ok.

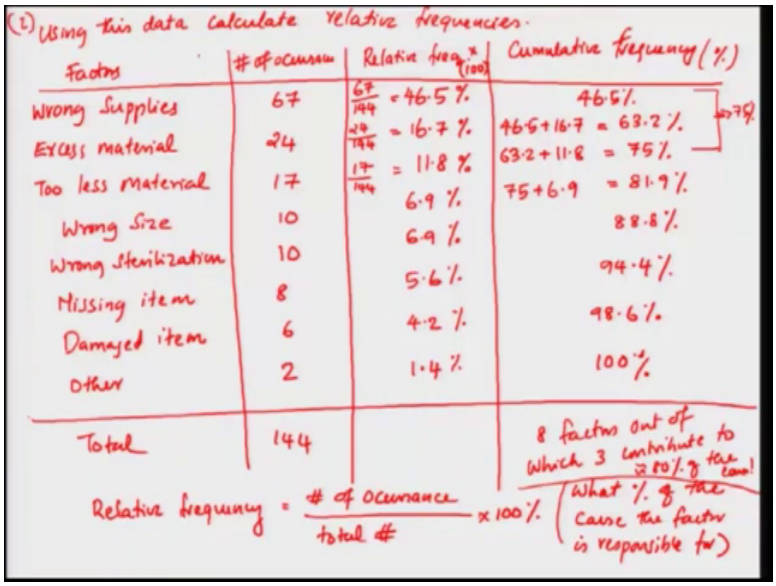
So, the data is let us look at what the data is? The data says less too less material ok, let us say here is the causes and not the causes let us say factors ok. Let this be the factors and this be the number of occurrence ok. So, too less materials, it results in 17 occurrences or it results in 17 occurrences of delay in time taken towards the delay in time taken ok.

So, then let us say wrong size is another factor; wrong size amounts to 10. Then, missing item accounts for 8; then, wrong supplies they account for 67. Then, we have let us say damaged item, it accounts for 6 of those. Then, we have excess material this accounts for 24 of those cases. Then, wrong sterilization, it accounts for 10; then, others this account for 2 ok. So, let us see how much it is? It will 10, 12, 20, 27, 34. So, this whole thing total of this accounts for 144 of this occurrences ok. So, you have 1 2 3 4 5 6 7 8; 8 items here ok.

So, now the question is you have 8 of them, when all of them are contributing towards the more time taken or this resulting in delay of or it is adding to the time taken to decide the treatment protocol in an emergency room of a hospital. In an emergency room you would rather like the patient to be treated quickly and in any delay could be you know life consuming. So, the question here is how do we focus on or which are the factors we need to focus on in this regard ok?

So, the first thing that we do is in a data like this; what we do is we arrange them in the descending order. So, the Step - 1 arrange data in the descending order ok. So, in this case, if we do that, then we have you know the data gets arranged in the form of Wrong supplies that will be 67 and the next is excess material that is 24, then Too less material that 17 ok; 17 like this all the way to others which is 2 ok. So, in the decreasing order of the count or the number of occurrences, you arrange the data in the descending order ok. I think that is very simple to do. So, all of you can do that.

(Refer Slide Time: 22:35)



Now, once we have this, the next thing we do is using this data calculate frequency relative frequencies. So, I will write this is Step - 2 ok. So, what we will do is I will write the in a tabular form and I will explain to you how the calculation was done. So, we will say is as the factors and we will have number of occurrence ok.

So, the first factor is as we said in the descending order wrong supplies ok, the number was 67; then, excess material the number was 24; then, too less material that is 17; then, wrong size 10; wrong sterilization 10, missing item 8; damaged item 6; other 2 ok. So, we sorted this in the descending order and we have what we call as the total column, they summed to 144 ok.

So, now the next thing what we what I told is calculate the relative frequency. So, let us call it as relative frequency alright. So, the relative frequency is calculated by the equation I will write it here. Relative frequency is equals number of occurrence divided by total number ok.

So, in this case then it will be 67 divided by 144 that would be 46.5. This will be 24 by 144 that will be 16.7. Next one will be 17 over 144 that will be 11.8. Next one will be 6.9 which will be 10 divided by 144. Next one will also be 6.9 10 divided by 144. Then, it will be 5.6 which will be 8 divided by 144. Then, 6 will be 4.2 which is 6 divided by 144 and 2 is 1.4 ok.

So, this 1.4 is these are all not relative frequency. This relative frequency multiplied by 100 ok. So, what I am trying to say here is these are the percentages; 100 percentage ok; so, this much percentage. So, what it actually says is the wrong supplies accounts for 46.5 percentage of the time. Why the patient is not gives? So, this tells you what percentage of the cause, the factor is responsible for. Where most of the people will be attributing the same thing to sterilization, you can see the sterilization only accounts for 6.9 percentage of the time when the patient is not given the right treatment; whereas, Wrong supplies is the major reason; one of the major reason ok.

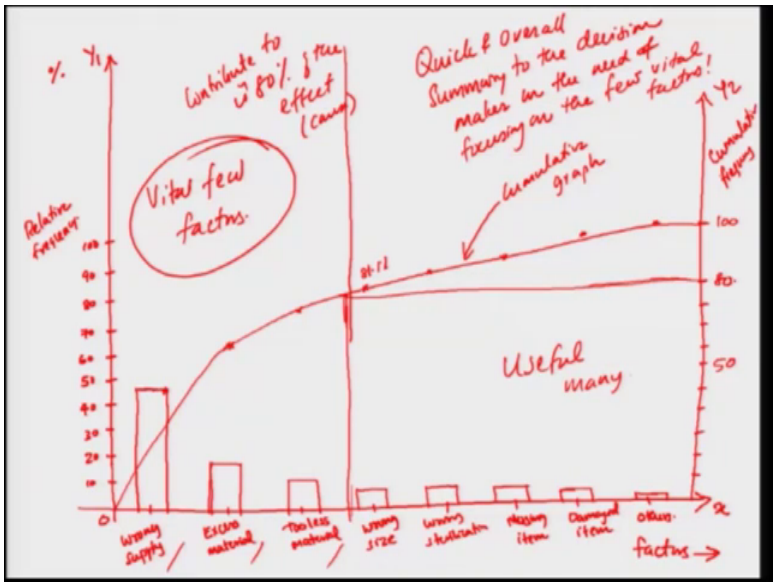
Then, the second one what we call as a cumulative frequency ok; cumulative frequency and what this does is this also is in percentage, but what it does is we keep on cumulating accumulating the sum. So, this will be first cumulative frequency will be 46.5 percent. The next one will be 46.5 plus 16.7 which gives you 63.2 percentage ok. Next V will be 63.2 plus 11.8 which will be 75 percentage. This will be 75plus 6.9 which will be 89.1 percentage. Next one will be 88.8 percentage which will be 81.9 plus 6.9. Then, 88.8 plus 5.6 will give you 94.4 percentage. Then, 94.4 plus 4.2 will give you 98.6 percentage; 98.6 plus 1.4 will give you 100 percentage ok.

So, ideally speaking when the cumulative frequency is done, the last value should some to 100 because all of these factors account for the 100 percentage. All these factors put together tell you that these 8 factors contribute to 100 percent of the reasons, why the treatment is delayed for the particular patient in the hospital.

Now, if you look into this, you can see that the cumulative frequency the first factor the first 2 factors account for 63 percent. The first 3 factors put together accounts for these 3 accounts for 75 percent of the time. Why these? Why the patient is not getting the treatment or the treatment is getting delayed in the emergency room. So, this is one of the reasons if you look into these there are 8 factors ok. So, there are 8 factors out of which 3 contribute to about 80 percent of the cause. So, 3 out of 8 ok; we can think about this about 20 percent right in slightly more than 20 percent, it does not matter.

But, what we are saying is the numbers are limited; they are not a major not a large chunk this much less than half of the reason ok. So, now, how do we this is our table. Table we can consider it as a diagram, but how do we draw it as a diagram ok.

(Refer Slide Time: 29:51)



So, we look into this the way to draw this as a diagram is this is kind of a tricky diagram. Usually, you have a x axis; you know y axis sorry my straight line drawing skills are very bad, but let us see where I can draw a straight line. Yes, I can 0. This is x; this is y ok. Let us call this as y 1 for the time being ok. Then, what this axis is trying to do is here we have the factors in the x axis, we will list the factors. We have 8 factors. So, we will write the 8 factors here 1 2 3 4 5 6 7 8 and then, you have your percentage here ok. So, then you have 10 20 30 40 50 60 70 80 90 100 ok. So, you have all the percentages going there that way right and these percentages will be we can think about as relative frequency ok.

So, and the factors here is the first one is the wrong supply ok, then the next one is excess second one was what excess material and then, the third one was too less material ok, then the next one is wrong size, then wrong sterilization, wrong size, wrong sterilization, then we have is missing item, damaged item and others; missing item, damaged item and others and the wrong supply, we can see there it is 46.5 percent is the relative frequency. So, we take these relative frequency values, these values and we plot and then we are plotting them ok.

So, if you look into that then we have 46.5; so, 40 somewhere here, 46.5 ok. We put a bar chart right here that is Wrong supply 46.5. Second one is the excess count 16.7. So, somewhere here we draw this ok. We can draw this beautifully using excel or anything

like that I am drawing this to demonstrate that is it. Next one is 11.8 ok. So, somewhere here ok, then we have is 6.9 two values ok. So, somewhere here ok, same way 6.9 two values. Then, we have is 5.6, then 4.2 and 1.4. So, 5.6 would be slightly less than this not too less, but slightly less. Then, we have is 4.1, 1.4; 4.2 will be still lesser than this not too much 1.2 is like a small bar here.

So, these ones. So, some people would write the values right here; some people would not right. So, this is the initial aspect of a Pareto diagram, but Pareto diagram is not over here because you typically draw 1 more axis right here which is what we call as the y 2 axis, where we do the cumulative percentage cumulative frequency ok. So, we have let us say same 10 20 30 40 50 60 70 80 90 100.

Now, assume that these 2 are of the same size can bad in doing this you typically do this at the same exact fashion. So, the first one is we will just kind of use this, but we have a second axis y 2 axis. So, the first one is the cumulative percentage was let us say 41 46.5 and the next one is 63.2. So, ideally speaking 63.2 is right here ok.

Then, the third one is 75. So, 75 somewhere here, you can see that this axis is wrong. Ideally, the cumulative percentage is come from here the cumulative comes from the left side same y 1 same, not same scale of y 1 ok. I have not drawn it properly you require a proper graph paper or graphing tool to do this. Then, 75, we can use excel to do this and the wrong size is if you look into this that is wrong size comes to 8 sorry too less material comes to 81.9 percent. So, the next one is 81.9 wrong size comes to 81.9, it comes somewhere here 81.9 ok.

Then, comes is 88.8. So, 8.8 is somewhere here ok, then comes is 94.4 ok. So, somewhere here 94.4 somewhere here ok, then is 98.6for this is somewhere here then, 100 is ok. Let me erase this 100 is somewhere here. The 100 should also this, 100 should also match right here ok. So, should look at 50 some place here. So, I am sure that they match. So, if you draw from 0 ok. If you draw a curve like this to this, this is the cumulative graph ok.

So, with this now, if you look into what we call as 80-20 percentage, so, let us look at somewhere where is this 80 percentage, ok. So, let us call this as the 80 prospect. So, let us draw the 80 ok. Actually it is not does not draw you do not draw it from there, you actually draw it from the cumulative. So, 80 comes here ok. So, comes and meets here if

you look into this, this portion, so, the first 3, 80. This is 81 percent actually. This is 81 point something percentage of that. So, this 80 comes and meets here.

So, these values these 3 ok, these are the vital few factors. So, these 3 factors wrong supply, excess material and too less material; these 3 contribute to the majority. So, these vital few factors contribute to approximately 80 percent of the effect or the cause ok. Then, this is what we call as the useful many ok. They are not vital, but they are useful; you can use them and you can understand them and you can identify you know why it is important and those kind of things down the (Refer Time: 39:03) ok.

Now, with this, what the major thing about this is that this graph is able to so this gives you a quick and overall summary to the decision maker, on the need of focusing the need of focusing on the few vital factors ok. So, this will allow you to understand why this thing is this graph, this particular why this particular what we call as our Pareto Diagram is an important graphical summary tool for the decision makers alright.

So, then we now get into the next tool which is called as a frequency distribution and what we will do is we will continue frequency distribution after a small break.

Thank you.