Advanced Green Manufacturing Systems

Prof. Deepu Philip
Dr. Amandeep Singh Oberoi
Department of Industrial & Management Engineering
Department of Mechanical Engineering
Indian Institute of Technology, Kanpur

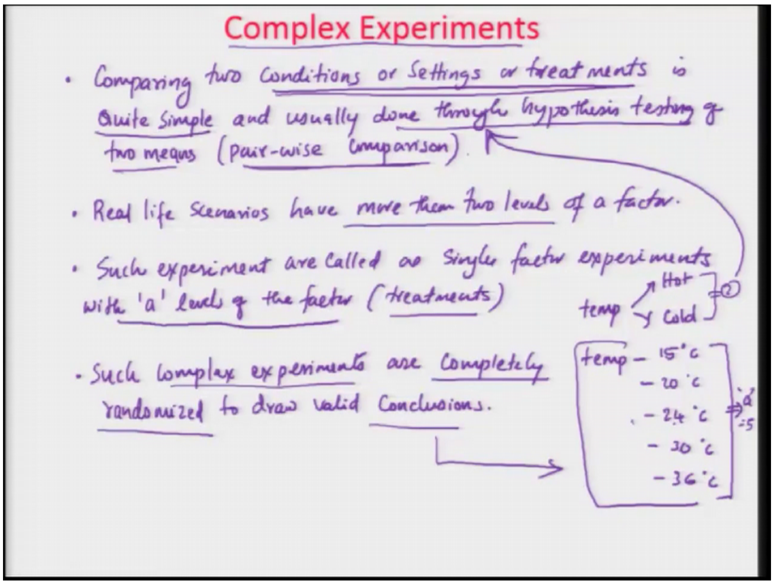
Lecture - 26
ANOVA – Part 1

Good afternoon students. Welcome to the course of Advanced Green Manufacturing Systems. And today, we are going to discuss new technique are or a special technique called ANOVA analysis of variance, which is in continuation with the experimental design concepts that we are covering in this course and I am Dr. Deepu Philip and I am from IIT Kanpur. So, if we look into today's, lecture it is titled ANOVA or what it stands for it is called single factor experiment analysis of variance ok. And this term ANOVA comes from analysis of variance. This is how you get ANOVA right.

(Refer Slide Time: 00:47)

So, whenever somebody says I am working, I am doing ANOVA; that means, what they are doing is there are doing, they are analysing the variance and today we are talking about the single factor experiment. How we focus on one factor and how the experiment is being done?

(Refer Slide Time: 01:19)



So, the first thing we need to study today is something called complex experiments right, not the simple experiments right. So, first thing is comparing two conditions, two conditions or settings or treatments comparing this is quite simple ok. So, if you want to compare two conditions or settings or treatments it is quite simple and usually done through done through hypothesis testing, hypothesis testing of two means or what we called as pair wise comparison ok.

So, this is a the; if you want to just compare two things, two settings or two conditions is that what you want to compare. This is, it is a reasonably a simple process and under the hypothesis testing most of the textbooks, statistics books gives you under the hypothesis testing of two means or the pair wise comparisons aspects will actually give you how to do this right. Real life scenarios have more than two levels of a factor two levels of a factor ok.

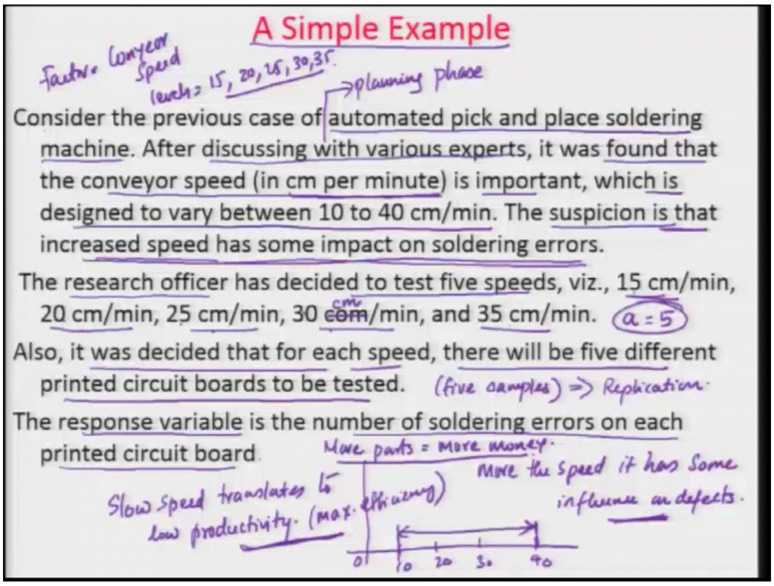
In real life, if you take one factor it is very rare that you will come across that there is only two levels of that factor. These always more than two levels of a factor and when you have more than that, then this becomes complex and you cannot really do with the simple approach. So, we called such experiments, such experiments are called as are called as single factor experiments, single factor experiments with a levels of the factor or what we called as treatments ok.

So, what we are saying here is when you have more than two levels then we say the factor has a levels of or a treatments. So, assume that you have something called a temperature as a factor. The two level will be hot and cold ok, but in real life it is not hot and cold people will talk about temperature varying from let us say, 15 degree Celsius, then you have 20 degree Celsius, 24 degree Celsius, 30 degree Celsius, 36 degree Celsius, something like this. So, these ones are the a levels multiple levels.

So, 1 2 3 4 5, so the value of a here is a is equal to 5, in this particular case. So, whereas, this is the two level, which you can do with this case ok, where you cannot do this with that you have to for this particular case, you have to use a another method right. So, such complex experiments are completely randomized; randomized to draw valid conclusions. So, what we are doing here is, the randomization is complete. You you randomize the entire experiment, complete when you have such large number of factors.

So, these complex experiments you how to do complete randomization, so that you can draw valid conclusions out of it right. So, that is the thing that we talk about the complex experiments.

(Refer Slide Time: 05:53)



Now, let us start with the simple example, let us discuss this whole thing as an with an example right and this example will help us to work through this single factor or analysis of variance with multiple levels and also figure out the mechanics behind the problem. consider the previous case of an automated pick and place soldering machine.

So, we were talking about how do, which machine we need to buy and what happened is the initial planning of the experiment that was discussions with various experts, after discussing with various experts, which is part of the experiment planning. So, this is the planning phase ok, it was found that the conveyor speed ok. So, we identified that the conveyor speed in centimetre per minute is very important. So, how quickly the conveyor of the machine runs in centimetres per minute ok is important and it typically varies between, this designed to vary between 10 to 40 centimetres per minute.

So, it is not one value, it can vary from 10 to 40, say if you think about it the, if you want to think about it as a graph, then you can think about it as 0 10 20 30 40 and the conveyor speed can be any value between this. The suspicion is that what you thing is that suspicion is that increased speed has some impact on the soldering errors. You think that more the speed, more the speed, it has some influence on defects. Let us say, you think that more the speed it has more defects then you; obviously, say then reduce the speed, why do we need to increase the speed.

But think about this slow speed translates to low productivity ok. So, you want to run the machine at the maximum efficiency, what we call it as max efficiency and this is one of the most important learnings that we do in green manufacturing is, probably running at the max efficiency is not the right thing to do, because you might not require to run it at that efficiency. Instead what we are doing is, we are trying to find out what is the right set of speed, where you can do is where you can minimise the errors and minimise the energy utilisation and those kinds of things.

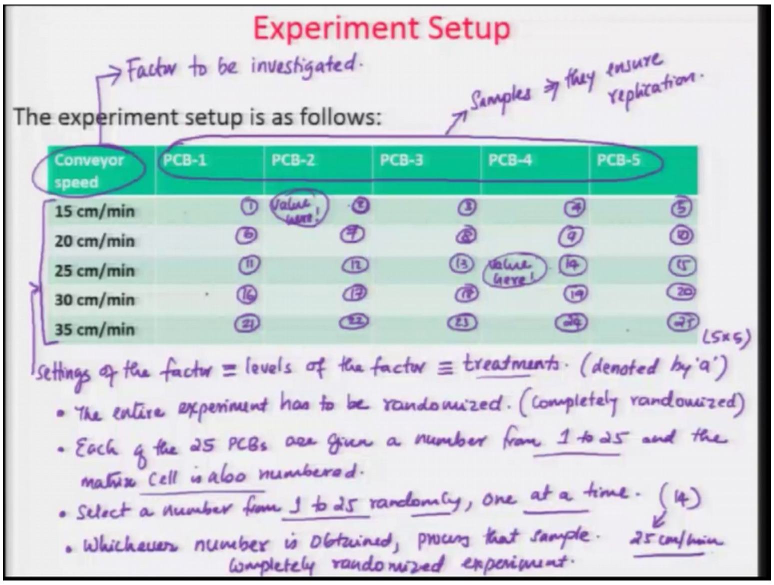
So, slow speed in normal scenario, normal production, you want to run the production as fast as you can, because you want to make more money by selling more parts so, which is, which the reason is more parts equal to more money. This is the productivity mantra right, you produce more, you make more money right, but what we are trying to talk about green manufacturing is more about not more, about producing that. So, this is kind of an practice problem for you to get this kind of concepts into your mind.

So, we suspect that there is suspicion that the increased speed somehow impacts the soldering errors or the soldering issues are somehow impacted due to the increased in speed. So, the research officer one person, who is doing the research has decided to test five speeds. So, he picked five speed he or she picked five speeds 15 centimetres per minute, 20 centimetres per minute, 25 centimetres per minute, 30 centimetres per minute, oh this is not 30 com, it is 30 centimetres per minute and 35 centimetres per minute.

So, you have 15, 20, 25, 30, 35. So, the here value of a is equal to 5 in this 1, 2, 3, 4, 5. So, you have five levels or five treatments for the one particular factor. So, the factor here is conveyor speed levels is 15 20 25 30 35. This can also be called as treatments the levels can also be called as treatments ok, also it was decided that it was decided that for each speed there will be five different printed circuit boards to be tested. So, for each one of the speed, we will take five different samples.

So, this is five samples will be tested ok, this is equivalent to your replication. You will replicate it five times for each speed and what is your response variable; response variable was decided as the number of soldering errors on each printed circuit board. Once you run at, then you what you are trying to find out is how many soldering errors are there, on each printed circuit board that is going to be your response variable fine. I hope you are clear with the experimentation part.

(Refer Slide Time: 11:05)



So, the experiment setup that we have is as follows right. So, you have the conveyor speed as I said earlier, this is the is the factor to be investigated. See you are investigating the factor called the conveyor speed and the conveyor speed has, these are the settings of the factor. It is also known as levels of the factor ok, it is also known as treatments. They are all one and the same right. So, we have five treatments in this, five treatments each at the speed of 15 20 25 30 35 as the speed of there. So, these are the five of this values denoted by a ok, the variable a is used to denote this and for each speed you have five P C Bs ok. These are your samples and they ensure replication.

Saying that I replicate it, five times for each experiment right. So, now, if you think about it, I can put numbers. So, there are five of them here and there are five of them here. So, I have 5 by 5 system. So, which is 25 total. So, I can do one thing, I can have this. So, the thing is number one; the entire experiment has to be a randomized has to be randomized, because as I said earlier, complex experiments has to be completely randomized ok. We can say completely randomized.

So, what do we do? We take each one of them and then we assign a number. So, let us take it as 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25. So, I have 25 printed circuit boards with me and each one of them, I will assign them a number from 1 to 5 1 to 25 right ok. So, what I am saying is each of the 25 P C Bs are given a number from 1 to 25 and the matrix cell is also numbered ok.

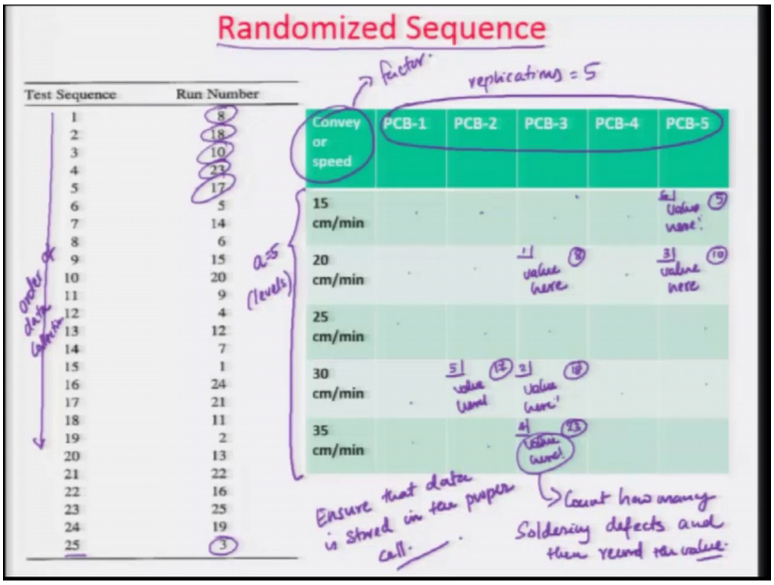
Then what do we do is, select a number from 1 to 25 randomly one at a time. So, what I am doing is first, I have 25 P C Bs with me, each one of the P C B I have numbered 1 to 25 and each matrix, I will also have number 1 to 25 in this case there 25 cells here right. So, they are also numbered, then I select a number from 1 to 25 randomly pick a number and one at a time and let us say I pick the number for an example is 14 ok. So, then what happens is I find out where is 14 in the matrix.

There is a 25 centimetres per minute and second is the column of P C B 4 right. So, then what we do is, whichever number comes in whichever number is obtained, process that sample. So, what do I do here is, this means if the number is 14; that means, I run the speed at 25 centimetres per minute and do the soldering process and whatever the error I get ok, then value here, whatever the number of errors, after the soldering process I whatever the value I get, I put the value here then I take another example say randomly pick the another number called 2. So, then 2 is right here, where the speed of 15 centimetre per minute and this is second P C B.

So, I take the P C B, which is labelled number 2 and then I run the machine with the conveyor speed of 15 centimetres per minute and then I find out how many soldering defects are there and I put the value right here ok. So, I repeat this process and then I probably would get number of 25 or something like that, then I will go, do it here, then I may get a number of 16 so, then I will come and do it here like this. So, I will not be doing them in an order, I will actually doing them in the random order.

So, that is why it is called as a completely randomized. So, this is called as a completely randomized experiment. All the 25 values are taken at random and that that that values are the numbers are filled in this process ok. I hope you guys understand, how this actually works.

(Refer Slide Time: 17:27)



So, let me give you an example, let me tell you how it actually looks like right. So, the conveyor speed I said, these are the is our a equal to 5 and here is your replications equals 5 here is your factor ok. Here is your levels etcetera and your test sequence is the this is the number order in which the there are 25 of the test sequence right here and this is the order in which you are do going to do the test ok; order of data collection ok.

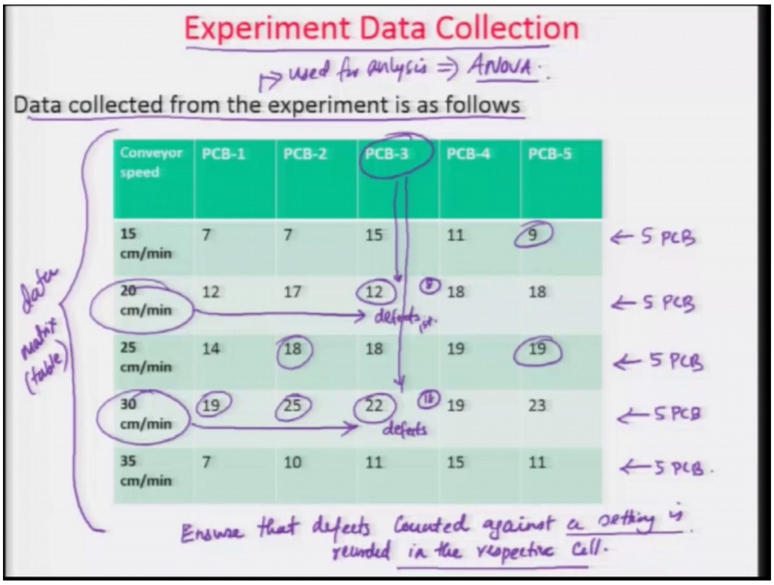
So, the number one is random number run number 8. So, 8 is 1 2 3 4 5 6 7 8. So, this is my 8 ok. So, I will take this one. So, run it at 20 centimetre per hour, use the P C B, this particular P C B, take the data, put value here, then the next one is 18 ok; so, 8 9 10 11 12 13 14 15 16 17 18. So, this is 18. So, then I have value here, then this 10. So, it is 8 9 10. So, here you have your 10. So, then you have your value here ok. So, depending then comes the value of 23. So, it is 18 19 20 21 22 23.

So, you got 23 right here. So, you have your value here, then you have is 17. So, 17 comes her so, that you puts the value here right. So, then comes what you so, you done with the seventeen then you have 5. So, here is your 5 so, then you go there value here like this. So, you keep on putting the value ok. So, you have some numbers whatever so, the value is the, count how many soldering defects and then record, record the value. So, then you put the value how many soldering defects, you got, you were put it right there.

So, this one and the you know that the last twenty fifth value is 3. So, the 3 will be right here, this will be your twenty fifth run or in another way to think about it is, this 8 will be your first run or in another way to think about it is I am just going to put the numbers right here. This is the value that you are going to collect as number 1. 18 is number 2 ok, then you have 10 is number 3, this is the third value that you collect, then is your fourth one is 23. So, this is your 4 ok, then 5 is 17, here is your fifth value, then the sixth value is 5 ok.

So, here is your sixth value. So, this is the order in which you cannot. So, either way you can label them, but ensure that when you do this ensure that data is stored in the proper cell ok. So, you have to ensure that they just do not write in wrongfully, just ensure that one way or other you store the value in the right place, wherever is the right location, you store the values right there ok. So, this you have to ensure that correct ok. Once this is done, this is what we called as the this entire order. This entire thing is what we called as the randomized sequence of the experimentation all right.

(Refer Slide Time: 21:21)



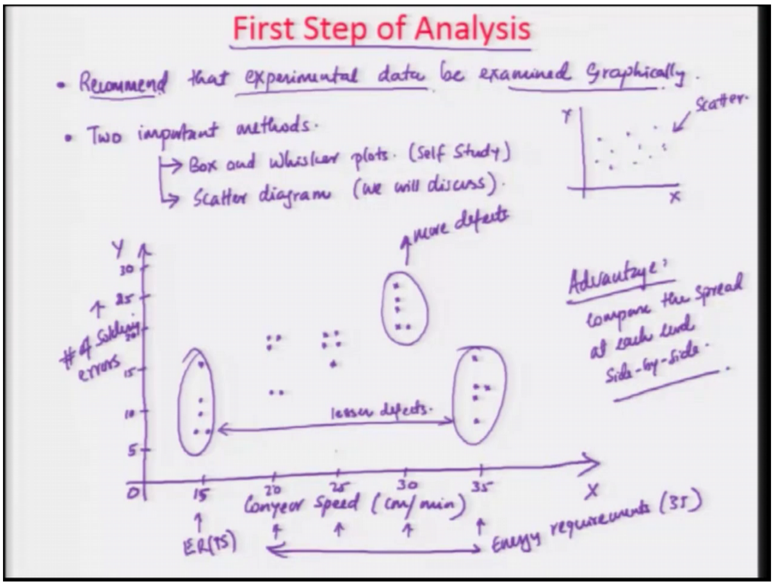
Now, that we have seen what it is let us move towards the experimental data collection. So, now, we see that we done the experiment and we completed it. So, if you look into this in the previous one this was the run order in this ok. So, in this case then we know that the first sequence was 8 ok. So, 8 means we went up to 5 plus 3, this was your 8, 8 was the column right here right. So, what happens is you did the experiment at 20 centimetre per minute and then you did using this particular P C B and that value is right here ok.

So, this means there are 12 defects that you count that this was your first sample, then the next one was 18 right. So, that came to, if you think about it, this was the 18 right. So, this is 18. So, this is what you did the second one. So, this particular value gain this particular, it has prove found running at a different speed. So, there is 5 P C Bs allotted for each one of them ok. There are 5 P C B here, then there is another 5 P C B here, there is another 5 P C B here, there is another 5 P C B here ok, there is another 5 P C B here. So, you require 25 P C Bs and there you found out that in this particular speed, you have 22 defects right.

And so, you keep on collecting data and this is the data that you record accordingly ok. All these data values that you put, there are the values that you recorded right. So, as I said earlier, ensure that defects counted counted against a setting, a setting is recorded in the respective cell ok. So, as of now we are saying that ok, whatever the data you collect you ensure that you record it in the respective cell and once you are done with these 25 data sets then you get this entire. So, this whole thing is what we called as the data matrix or table data table some people called as table, some people called as matrix. It does not matters.

So, this whole thing is the data table that we are going, we are talking about and it is on this, this data collected on this. This one is used for analysis and what is our analysis ANOVA in this case ok. So, we use this data matrix to do analysis of various for ANOVA in this regard ok.

(Refer Slide Time: 24:19)



So, then we get into what we called as the first step of analysis ok, what is the first step in this? Recommend that experimental data be examined graphically or using a graph find a way to examine the experimental data, this is my recommendation this regard.

Whichever data you collect, make sure that you know you use graphical methods to examine. Two important methods, I would suggest two important methods in this regard, number one; box and whisker plots. I have already covered this in another lectures, in other courses, but I would recommend; so, I will suggest you to do self-study in this regard or look into the lectures that I have already made on box and whisker plots, then the other one is scatter diagram ok. Everybody has learnt, but I will we will discuss now, everybody has learnt what is the scatter diagram and it is usually between one X and a Y ok.

So, we are all seen that this is X this is Y and we have some points like this and we say this is a scatter ok. In our case we are actually going to do in a little bit of a different fashion. So, I am going to explain to you what is one example of this. So, this one what typically, in ANOVA the way we do this is something like this here is your X is your Y and X is your what were we counting we were counting the number of defects right. So, we has said the number of soldering errors is what we are considering. So, we will say number of soldering errors ok.

That is on your Y axis and here is the conveyor speed in centimetres per minute ok. So, you had what? You had conveyor speed of what was the conveyor speeds we are looking at 15 20 25 30 35. So, we have something like 15, we have 20 25 30 35. So, these are the four conveyor speeds that we have and we are going to take the value 0 and that we are going to take the values from this side is different and we draw it as 5 10 15 20 25 30 like this.

So, then with that in mind, so for the levels values of 15 ok, if we think about it the first values are we go back for 15 it is, 7 7 15 11 9 right. So, to draw that what we do is we go here 5. So, let us say this is 7 ok, you have 7 and 7 ok, 7 7, then the other values is 15 and 11. So, the another value is 15, third value is 11 and the last value is 9 somewhere here. So, this can refer diagram that we just draw right here or we drew right here is the scatter ok, the scatter for the particular conveyor speed of 15. Similarly, we look at the conveyor speed of so, these values were draw.

Now, will said look at 20. 20 was 12 17 12 18 18 right. So, 12 17 12 18 18 is what we are interested in drawing ok. So, then let us go back. So, this is 10 ok, maybe 12 is right here, 1 12 then 12, then comes you have is 17 15 16 17. So, this is 17 then you have is 12. So, come back and draw a 12 right here, then you have is 18 and 18 right. So, this is 18 and another 18 right.

So, you get a diagram like this, which is the another scatter diagram for 20, then we go off to 25. So, we look at 25 14 18 18 19 19 right, so 14 18 18 19 19. So, let us go back 25, the first one is 14, so slightly below 15. So, we draw here. So, that is 14 then we have is two 18s. So, we go up right here is 18. So, two 18s and then you have two 19s.

So, slightly below 20 so, you do two 19, so you get your five values therefore, 20 25 then 20 30 is 19 25 22 19 23 right. So, then let us go back, for 30 we have first value was 19.

So, right here 19 the next value was 25. So, we go off here is 25, then the third value was 22, so between this one. So, this was 22 then there is 19. So, right here is 19 ok, we put one more 19 and then there is 23. So, here is 23 between 22 and 23.

So, this gives you the scatter for the 30 and then the last one is 35, which is 7 10 11 15 11 right. So, then we go back to this and 35 and then we start with the value. So, that is 7 ok. So, somewhere here is 7 ok, 7, then you have is 10. So, you get 10 right here, then you have is 11 slightly above that right, then you have is 15. So, this is 15 and then you have is 11. So, you have one more diagram like this. So, now, we can see that each one of them, this is the, this is the way, each one of them is spread out right.

, so, this is for the speed of conveyor belt of 15, this is for 20 25 30 35. So, the advantage is, the main advantage is compare the spread at each level side by side or simultaneously. So, you are basically saying that by doing this kind of a diagram, we can actually compare them. So, we can say that 15 is probably the lesser number of defects, but also at a high speed of 35 also, these are comparable right, but the conveyor speed of 30 consistently shows that this is more defects and here you have lesser defects ok.

Now, for a green manufacturing process you can decide there is a speed of 15 and this is speed of 35; so, here is energy requirements for 35. Here, you have your energy requirements for 15 ok, compare these and see which one is better and those kind of things. So, this is how you actually work the problem, but you whatever it is, you have to still compare using Anova to figure out whether this is better or bad or worst or whatever you want to call right fine.

Thank you very much for your patient hearing and the wish you all the best.

Thank you.