**Engineering Metrology**
Prof. J. Ramkumar
Dr. Amandeep Singh Oberoi
Department of Mechanical Engineering & Design Programme
Department of Industrial & Production Engineering
Indian Institute of Technology, Kanpur
National Institute Of Technology, Jalandhar

Lecture -37
Data and scales in measurements

Good morning welcome back to the course Engineering Metrology. In this module, we discussing the statistics in metrology. So, in this lecture I take what is data And what are the scales of measurement. These are not the scales the physical scales, these are actually statistical scales or the scales how the data is organised ok.

(Refer Slide Time: 00:41)
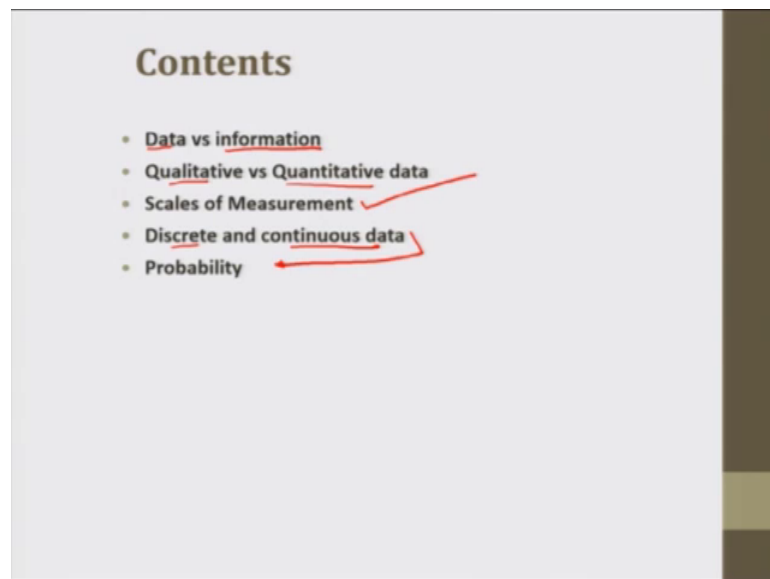
So, the contents would move like this. I will try to compare the data and information these two words are used interchangeably and multiple times you might have heard we have we have this information we have this data, but there is a difference between these two words I will just discussed that. And then we will discuss about the types of data qualitative versus quantitative data.

Then we will discuss the scales of measurement, then I will discuss about discrete and continuous data and how these are used in probability and probability distributions. Before that we will just look into what is probability you people also know what is

probability, but how the frequency of probability distributions are designed and why and what can we extract from that what can what influence can be taken out of the fitting the data the specific data that we have obtained from measurements from the physical measurements how can we fit that into the specific distribution on which distribution has which kind of applications these things we will see.

(Refer Slide Time: 01:48)
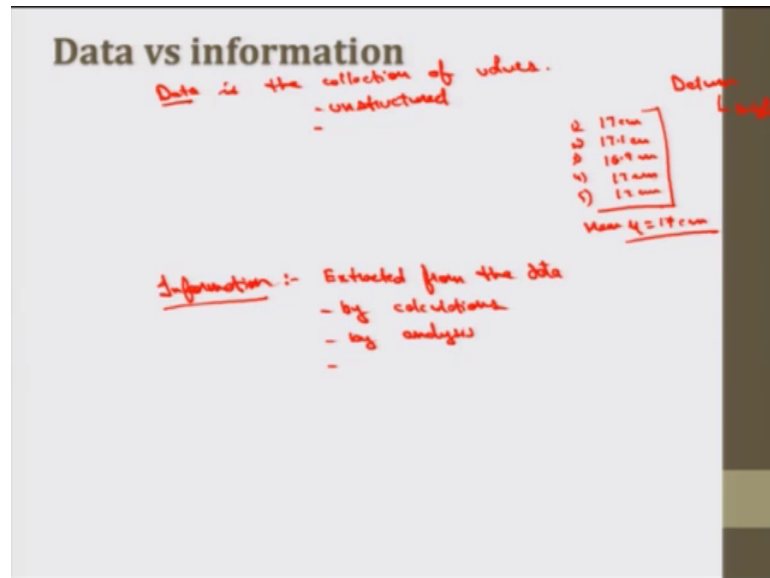
So, first of all let us see what is data? Data is the set of values those are collected by some measurement by some observation or maybe by from the secondary information as I have discussed that the two kinds of data primary data and secondary data. Primary data is that we collect from the basic source from the actually we do the experiments and then we find the value. We observe the value then we record it that is primary data whatever in laboratory demonstration. Whatever we have been doing that was primary data. Secondary data is the data that is already record recorded for instance the standards the specimen standards those are there those are secondary data. So, data is the I will put it definition data is the collection of values.

So, this has come from the Greek word datum. Datum means which when I use the word Datum it means the single value if I say height of the stylist the height of the stylist is about 15 to 17 I will put it a 17 centimetres this is 17 centimetre. This is one value and if I take multiple stylisis and take the height of that that might become data when I manufacture that ok. So, this is one Datum this is collection of values and if I keep on

collecting the values, if I since I have the values like 17 centimetre 17.1 centimetre 16.9 centimetre 17 centimetre; this is just data. And if I talk about a data broadly like even out of the scope of our engineering metrology, data is a very wide or brought down that is used for the numbers. The numbers or maybe actually the two types of data qualitative and quantitative data. So, generally data is it may be numbers or it may be some attributes ok.

So, they also terms like big data like a industry there is a lot of data regarding like in the whole life cycle or using the whole I would say supply chain cycle of the industry from the value of raw material to the finish good we do in metrological applications are always there ok. When we have to get the raw material, we check that whether the raw material that we have received is that acceptable or not. We have to do some measurements we have to do some calculation sometime ok. So, h it point data is collected. So, big data is whole supply chain who is the vendor, what are the prices, what are the market trends and what would the, what would customer like to pay. All see data is available in all the kinds of face of the manufacturing.

But it is not in a structured way it is unstructured data. So, data is generally unstructured. So, in scientific research data is collected by a huge range of organisation institutions and that is used to extract some information. Now what is information? Information is when we try to extract something from the data and a like some evaluation some calculations are done and data is a little structure or organised for instance this is the data all these points 1 2 3 4 and 5 this is data the height. If I take mean value mean mu appreciate the mean would be about 17 centimetre only.

This is information mean value, median, we have the mil the mean value of the n number of components level of the height of n number of components is this is 17 centimetre ok. This is information is extracted from the data by calculations or by analysis ok. There is a netted term that I will not discuss more that is the knowledge. Knowledge is actually personal when we have a lot of information like if I know if I think the generally used what is do you have the knowledge how to operate this instrument it is knowledge actually it is not the information I can I have the information the procedure to use this.

The knowledge is the connection of knowledge is actually gathered all in your brain. When I have hold the instrument I have the knowledge of all the things how to hold that

what to hold in my left hand how to use my right hand right hand can be used to work in more precised way, it for precision movements and how to hold it, where should I how should the instrument look where, how should I read that. This all this things when those are connected the final application from the information that we can use finally, is actually knowledge ok.

Knowledge is personal. So, the instructors, the practitioner, the lecturers the management people they have different kind of knowledge sets. So, that interconnection of the knowledge swears or feels make them expert in something ok.

So, this is data and information this is a general just general on overview because these two terms were being used interchangeably multiple times and just to make it a little clearer, I am trying to get a little. So, gathering data can be accomplished through a primary source on the secondary source data analysis methodologies can vary depending upon different kinds of availability environment and maybe data triangulation data percolation or the few methods and we can we basically we need to discuss all this.

(Refer Slide Time: 09:11)

I will move to the next part qualitative versus quantitative data. The terms in self tell something about that. Qualitative is the qualitative data is the one which just tells about the quality it is about the attributes ok. I can say good or bad; the quality of the pen is good or bad acceptable or non acceptable if I am using the two terms good bad

acceptable not acceptable yes no, true false whatever that is ok. So, these terms this kind this kind of data is (Refer Time: 09:47) data I will discuss that as well.

So, qualitative terms; qualitative is when we use attributes ok. I will just differentiate it qualitative and quantitative ok. When I discuss about the height of some measurement some measurement height of that or some length I can say this is long, this is small, this is medium three categories ok. This attributes can be long medium and small it can also be if I use two terms long and small medium is in between. So, anything like colour can be there colour dark or light colour dark or pale dark or light ok. Then even if you say colour is yellow blue orange this is also a kind of qualitative data.

But if I give the values if I give the values when instance I am talking about the pencil. Let me say the height of the pencil if it is 17 centimetre it is long then 10 centimetre is medium 5 centimetre is small. So, these if I measure the exact heights this is these are actually numbers. These are actually numbers and these are measured Attributes are not measured these are observed 17 10 5 it could be 10 11 12 13 14 12.5 13 5 any number could come ok. The thing is that I am trying to explain is that we can even put the quantity data into bins to put the quailitative we have the qualitative information ok. I can see I can divided into multiple parts here ok. I can say from 15 and above it is long that is 15 and above 8, it is medium 8t cm and less than 8 is small.

So, actually medium is between 15 and 8 more than medium is long less than 8 is small. So, I have put the quantitative data into bin as well. So, the role of or the research that could be done in qualitative data is just exploration; this is big small, this is these are the things available to quantity is high low the temperature is cold cool or hot I can just explore something.

But we cannot apply very powerful statistical tools here. However, the statistical tools available for the qualitative data as well because in certain measurements like in go no-go gauges go no-go gauges are actually also derive from the quantitative data only because we have the value for the go and the no-go ends ok. But sometimes the data is quantitative data is not possible to attain ok, but then qualitative data is also to be worked on we can see the goodness of fit using chi squared test that will discuss, but the research is generally exploratory exploration ok. Here we can give some inference some inference some conclusion we can give some value.

This is the control limit above this value we cannot accept below this value we cannot accept. So, we can design something based on the quantitative data. So, quantitative data is always preferable if possible depending upon the cost the time if it is possible the cost the cost of the getting data primary data is too high.

We can even rely upon the qualitative data if that is acceptable if that that can be worked on if the time to get the data is too long then also be sometime we have to rely upon the quality data for instance the for you think about the data for the global warming ok.

If I think about the temperature in last hundred years the temperature of our globe for last hundred years and different points in our on the in different in different countries that is the quantitative data. Qualitative information or the information that can be extracted is the temperature rise can be the global phenomena global temperature rise is there global warming is here this is one thing ok. the quantitative data is the exact temperature exploration in the qualitative way I would say I can say is just hot cold if I talk about the temperature of my city Kanpur here.

The temperature is hot cold what exactly is the temperature. If I have to do some measurements using an instrument which is temperature sensitive can I do it in the open environment or not. For that I need to do know the exact temperature of the environment ok. So, for solar panels also like what is the, what amount of heat is hitting the solar panel for the p v panels of the of solar set solar plant. So, that also is that can be quantified as well. So, these are the few examples ok. So, next point I can say here that this is generally known statistical. Generally this can be used in statistics. As I just said I will very a very few tools are available or not tools not available actually could be applied to get some information from the qualitative data.

But yes if the data is quantitative we can apply statistics to get different kind of information and we can also conclude something ok. The sample here in the qualitative data is a non representative in the quantitative data the sample is representative.

So, this qualitative data is actually used for the initial understanding initial understanding. And from the initial understanding if we need the data has to be quantified then we can do the further experiments to have the detailed information or we have to detail have the detail understanding or the further course of action.

(Refer Slide Time: 17:35)
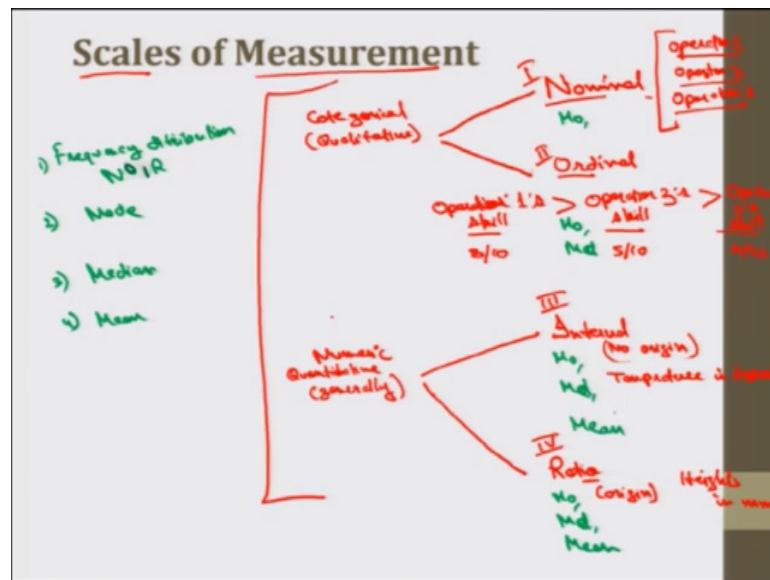
Next is the scales of measurement scales of measurement of very important. When I talk about data qualitative or quantitative. the qualitative data is also known as sometime it is categorical data actually yeah categorical data and this is numeric data this is actually already mentioned in 0.1 here. So, when the data is categorical or qualitative it can be of two forms if we these are stage is nominal or ordinal nominal data is nothing, but if the word is name here the word is nominal it comes for the word nom it can be we can just remember it is just the name ok. Like I say I have five samples I just number this sample number 1 sample number 2 number 3 number 4 number 5. I have just put the names what do these names represent it just helps me to recognise which is sample number 4.

But is it possible can I say that sample number 5 is better than sample number 4 without any other information can I say that sample number 5 is better than sample number 1or can I draw any this kind of information from this no.

So, this is just the name example here maybe I can say operator now operator 1 operator 2 operator 3. This is actually the operator who is doing the measurements. For instance if my factory is running for the whole day around for 24 hours and the workers were working on one machine and 8 hour shift is there. So, three workers will be working if a machine is working for 24 hours completely.

So, I have to name the three operators operator 1 operator 2 operator 3 ok. If I see that there is some trend like the quality that is operator 1 is giving is a little lesser than what

the operator 2 is giving it and the operator 3 lies between operator 1and operator 2 that can be ordered as well. That kind of data is known as ordinal data.

Now if I say operator ones this is actually the name operator ones I am telling operator ones skill is better than operator twos skill operator threes skill as I said and it is better than operator twos skill in working on a specific machine ok. Now I have put operative ones operator one is a noun here operator one like Amandeeps Ramkumars some.

Some you can they used some name Siddiquis Shubhams we can use some name this persons skill. So, operator ones skill level can be attribute or the skill level can be further we can quantify that skill level can be put in a on a skill maybe richer skill 125 skill were they have certain ways to do that as well, but I am just talking about that operator 1 operator 2 operator 3 were just names. Now this is the order for the skill. This kind of data is known as ordinal data ok. Like I said before in the qualitative data as well quantitative data can also be ordinal. In ordinal data for qualitative example I can say long medium small.

This is order long is greater than medium is greater than smaller ok. And in quantitative data with just have the information 17 centimetres is greater than 10 centimetres which is greater than 5 centimetres. Now this is the kind of scales nominal scale ordinal scale ordinal scale is now in nominal scale we cannot compare we cannot say that operator 2 is better than operator 1 whatever that is ok, but in ordinal scale yes we know operator 1 if I rate the skill from 1 to 10, let me say his skill level is 8t out of 10 this operators skill level is 5 out of 10 and this is 2 out of 10.

So, there is a order associated with it. So, this is ordinal scale. We use this scale in our numerical problem as well where is the ordinal scale working here and next kind of scale are the metric scales for the quantitative data. Now next I have kind of scale is interval scale. And next to that the final kind of scale is the ratio scale. These two kinds of scales are generally quantitative ok. So, the difference between the interval and ratio scale is that. So, actually there is a order associated where this is the first kind this is a second kind this is the third this is the fourth as we are moving from the first to fourth the extent of information is increasing.

We can more apply the statistical tools we can use it in a better way to draw some influence out of this. So, interval scale is a quantitative scale where the order or whatever the, with just had the order operator 1operator 2 operator 3 ok.

Now if I have the specific value for that for instances 8 5 2 was the value this is an interval scale. So, it is the classification or order or distance, but it has no origin it is important it has no origin. Here we have from the word the letter o we can say it has a specific origin. The example for the interval scale could be temperature.

We call temperature in degrees degree Fahrenheit degree centigrade's temperature in degrees. So, can I say that if the temperature is 25 degree for one body and 20 degree for one another body? If I put them together, can I say the total temperature is 45 degree? No. There is a big model with thermal model associated with it if there are two bodies; what is the material of these bodies, what is the surface total surface area of contact, what is the thermal conductivity of two bodies. So, what time would it take for the temperature to come to the same value at both these ends and at it is centre depending on the environment temperature.

There is a big model associated with that we can do that ok, but in general I will say the temperatures are not directly additional we cannot divide 100 degrees by 25 degrees directly ok, but ratio skill is one when we have the definite 0. We did not have the definite 0 for the temperature. We have the definite zero in the ratio skill it the example can be the multiple examples here they can be the example can be any lengths or heights in millimetres. So, or heights age then weight these all are in ratio skill we can say they if I having 5 kilos of apples if I divided into 5 parts each one would get 1 kilo of apples.

Because it is a ratio scale I can even say if there is a length if there is a there is a stick of 1 metre I need two sticks or three sticks of 33 centimetres or may be let me say and it is four sticks of 25 centimetre I can cut this into four parts. In other way 100 centimetres by 25 centimetres is 4 it is exactly the ratio is here ok. That is why it is known as the ratio scale, but we cannot say if a temperature is 25 degree I will divide it into two parts 12.5 degree to 12.5 degree 2 this is senseless ok. So, that was interval sales scales this is ratio scale whatever measurement the linen measurements

We are doing are mostly in ratio scale. So, where do we apply these? We will discuss about the probability distributions. In probability distributions we will use interval and

ratio scales mostly I will just use the nominal and ordinal skills to discuss about the few [stat/statistical] statistical tests, but most of the times we will work on the numeric data.

So, this is numeric data as I said quantitative data is numeric. Ok about scales we can talk for multiple hours you can ask the questions in the in queries and I will try my best to get all of them in the best possible way, but we have to move further for after this information on the kinds of skills.

I like to discuss further the two kinds of data which are discrete and continuous data. So, about these four kinds of scales what kind of statistical parameters can we use the order of values is known in ordinal interval and ratio data. In nominal data we do not have any order, but we can have frequency distribution for all the four constitute I can put here the frequency distribution this can be for an for all this number 2 what kind of the statistic can we use here the mode for nominal data mode is the maximum value for nominal data we can use mode, but not median because we cannot compare them ok.

We can have mode here ok. If I say m o mode is possible here mode is possible here and here. So, if I say I will put it here mode median mean. So, mode is possible in all these median. Median is the middle value we cannot have median in the nominal data, but in ordinal interval and ratio skills we can have the median middle value a like in order also we can have the middle value, but we cannot sum the order of and you can average that is not possible mean is not possible.

So, we can have median I will put median here before median I will put m d median and mean can be calculated only in the numeric data. So, we can multiply and divide in case of ratio skill it has a 20, but yes we can add and subtract some sometimes in the interval data as well, but multiply is not at all possible.

So, this was about the skills of measurement. So, next I like to take a little break here I will move to the next part of the lecture we will discuss about discrete and continuous data and then we will move to the probability distributions. I will not like to lengthen this lecture a lot. So, let us meet in the next part of the lecture.

Thank you.