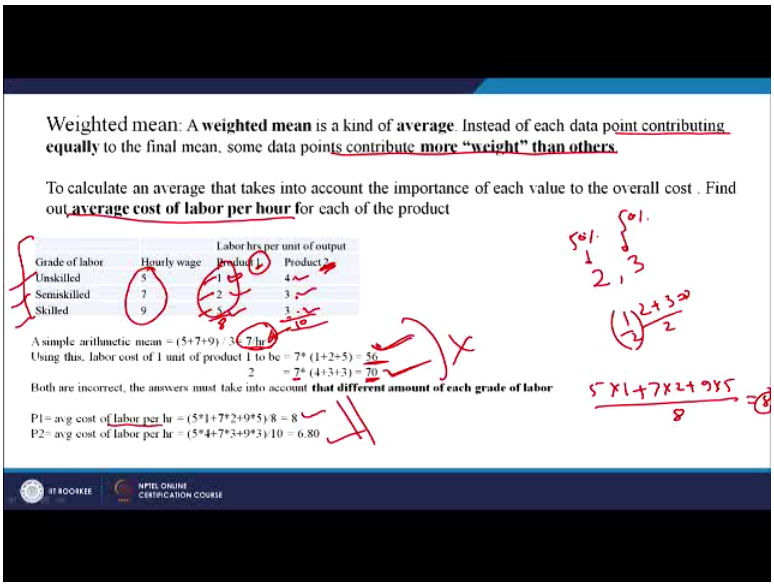
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology – Roorkee

Lecture – 06
Measures of Central Tendency – Part II

Good morning everyone, I welcome you all in this session. As you are aware in previous session we were discussing about how to calculate mean of grouped as well as ungrouped data. We have seen method wherein we assigned code to different class intervals and that is how we calculated mean in a better and faster manner. So let us look at another mean is called weighted mean.

(Refer Slide Time: 00:57)



Now, the difference between weighted mean and simple mean is that in case of weighted mean we assigned different weights to different data points while in case of mean we assign equal weightage. Just to give you an example let's say you are calculating mean of these two numbers 2 and 3 so the mean would be 2 + 3 / 2. So we are giving 50 percent weightage to this and 50 percent weightage to this, right.

This, that is what one by two means right? Now, if you use weighted mean then it is it is a kind of average where each data point is contributing differently not equally, right? Some of the data points would contribute more than others. Let us look at this example; there is an example in which there are three types of labors, unskilled, semi-skilled and skilled workers right or skilled

laborers. Now they are making two products product 1, product 2. So the labour hour, per unit of output is 1.

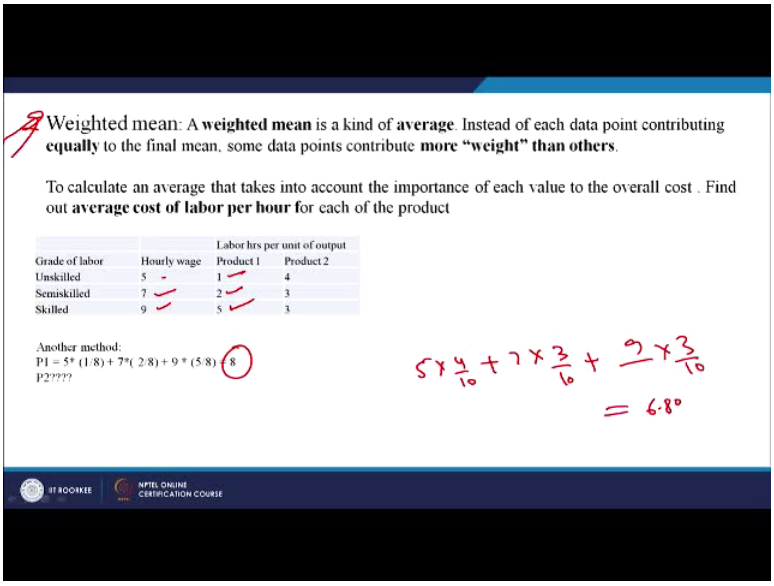
So to make one product or 1 unit of product 1,the labour requirement is 1 hour as far as unskilled labour is concerned, 2 for Semi skilled and 5 for skilled. Similarly to make one unit of product 2 the labour hours requirement is 4 hours for unskilled, 3 for Semi skilled, and 3 for skilled. Now the question is you want to calculate the average cost of labour per hour for each of these two products right? What is the question? The question is how to find average cost of labour per hour?

Now if you go by the method which we have seen earlier then you will simply calculate average of these 3 which is 7. We will say that the average cost of labour per hour is 7 but it not a right answer. Why this is not right answer? If we say this is the answer 7 hours then for making product one how many hours you are requiring? Total hours: 8 hours; 1 + 2 + 5 so 8 hours.

So, total hours required to make product to make one unit of product one is 8. The labour cost is Rs 7 per hour so the total cost would be 7 multiply 8 is equal to 56. Similarly for product 2, total hours required for product 2 is 4 plus 3 plus 3 equal to 10 right? 10 multiplied by 7. So, would we say that the labor cost of one unit of product one is 56 and for product 2 is 70? No, this two are wrong answers, why? Because, we have given equal weightages to all these three hourly wages, which we should not have done.

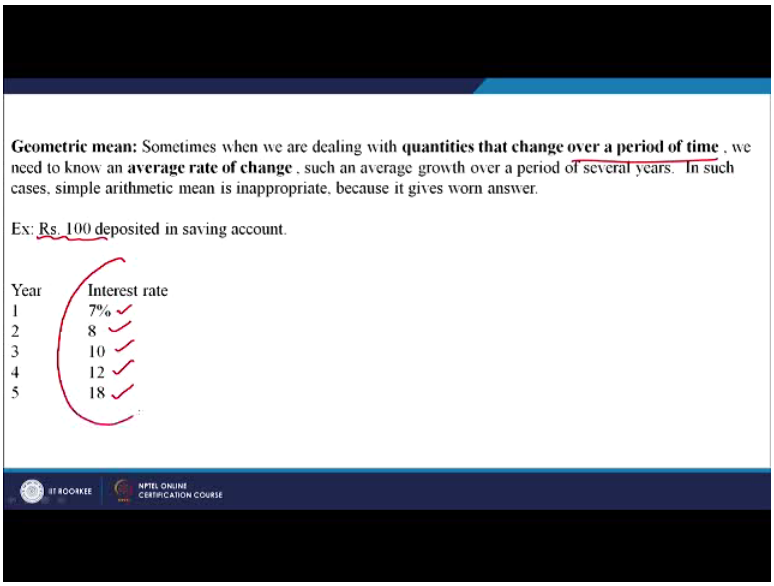
So there is a better method available and a correct method available. So let us look at what is that method? We know that the total hours required to make 1 unit of product 1 is 8. This is 8, summation of all this is 8. So what will do 5, 5 * 1, 5 * 1 + 27 * 2 + 9 * 5 divided by total hours and this becomes 8. So, average cost of labour per hour is 8 rupees. Similarly for product 2, how would you calculate? The total hours required for product 2 is what 10, right? So 5 * 4 + 7 * 3 + 9 * 3 / 10 divided by 6, so this is the correct answer. This is known as weighted mean.

(Refer Slide Time: 05:48)



Now, there is a different method of calculating same weighted mean what you can do is, this is 5 * (1 / 8) total number of hours 7 * (2 / 8) plus 9 * (5 / 8). So the answer is 8. How would you do it for product 2? Just think for a while is 5 * 4 / 10 + 7 * 3 / 10 + 9 * 3 / 10. The same answer you will get this is 6.80, right. So either you calculate by this method or by this method okay.

(Refer Slide Time: 06:53)

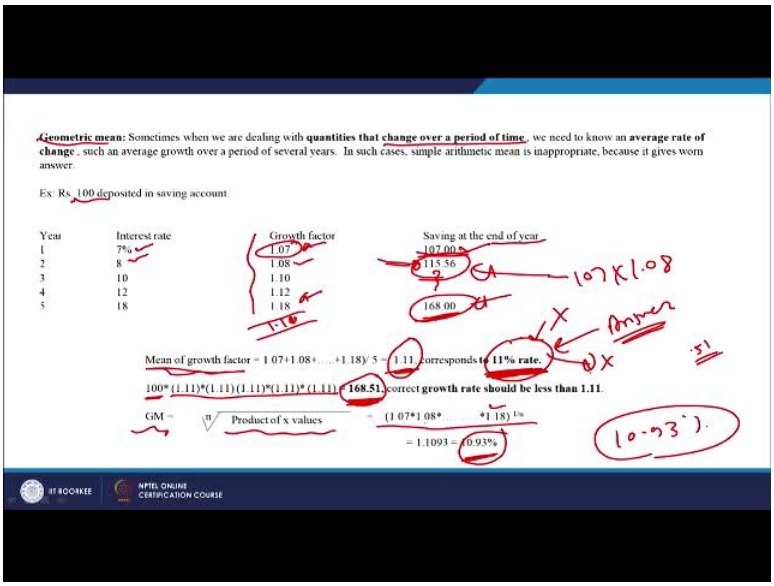


Now let us move on to another type of mean is called geometric mean. So far we have seen Arithmetic mean and weighted mean, right. Now there are situations where quantities change over a period of time. And when the quantities change you do not know what exactly the average would be? Of course you can calculate average, but would that be correct? Let us take an

example. So let's say, you have deposited hundred rupees in a saving bank account and you are getting 7% interest rate in the first year.

Now after 1 year the bank has increased interest rate to 8% then 10% 12% and 18%. Now we want to find out what is the average interest rate? or what is mean interest r? So let us solve this question.

(Refer Slide Time: 08:00)



So we know that the interest rate is 7% in the first year so growth factor would be 1.07. Growth factor means what? One rupee would become 1.07 rupees right. So hundred rupees would become 107? Savings at the end of 1st year is this. Now the next year interest rate is 8% so growth factor would be 1.08. So this would be the savings at the end of second year. How did you get this number? This is 107 * 1.08. Then you will get 115 and so on.

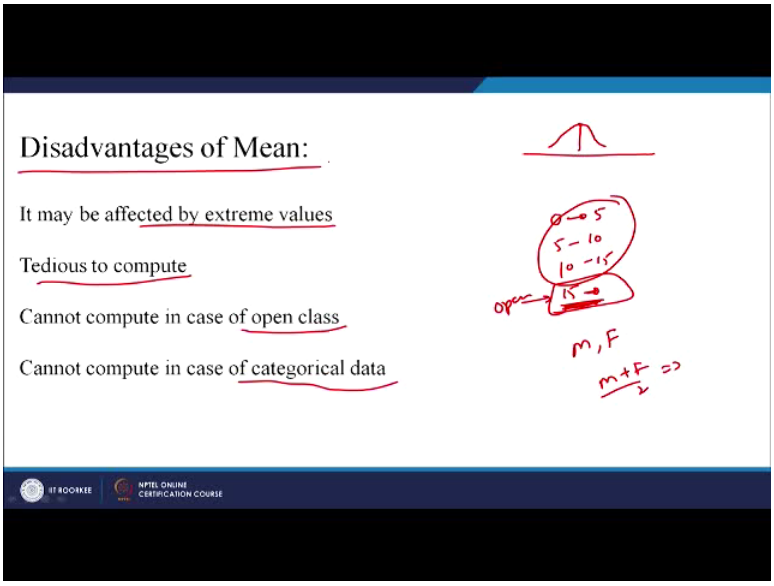
You just calculate remaining values. So, finally you will get 168 rupees at the end of 5th year and initially how much you deposited 100 rupees. Now if you take the average of these growth factor the average becomes 1.11. This is the one it means 11% because 1 Rupee will become 1.11 rupees and 100 rupees would become 111 rupees so rate is 11%. So we will say that the mean growth factor is 11%. To take this is answer that if you take for the time being this is your final answer and your mean growth factor is 11%.

Now if you take hundred rupees initially your interest rate is 11% so you should get 168.5 at the end 5th year. But here how much we are getting 168, so we are getting 0.15 rupee less. It means this answer is not a correct answer. So there is something called Geometric mean which will take care of change in quantity over a period of time. One more inference is since you should get 168.51 rupees but you are getting 168.

It means the mean growth rate has to be less than 11%. So what is that value? How much less it is. So for that, calculate Geometric mean which is an under root of product of X values. You can write this in this form. So, 1.07 * 1.08 and all other values growth factors. So 1.18 to the power 1 by N. N is nothing but five right. There are five values. So, this becomes 10.93. So, the interest rate is 10.93 it is not 11%.

If you use this particular method for calculating interest rate it would be a wrong answer, right. So, the correct answer is 10.93 is the interest rate. This is, this answer is wrong answer, ok. So we next move into next slide.

(Refer Slide Time: 12:31)



So we have seen weighted mean, Arithmetic mean and geometric mean. Now whenever what is mean actually? What, why, what, we why we are doing all these things actually because we are measuring Central tendency, right, isn't it? We are measuring debt point where the distribution

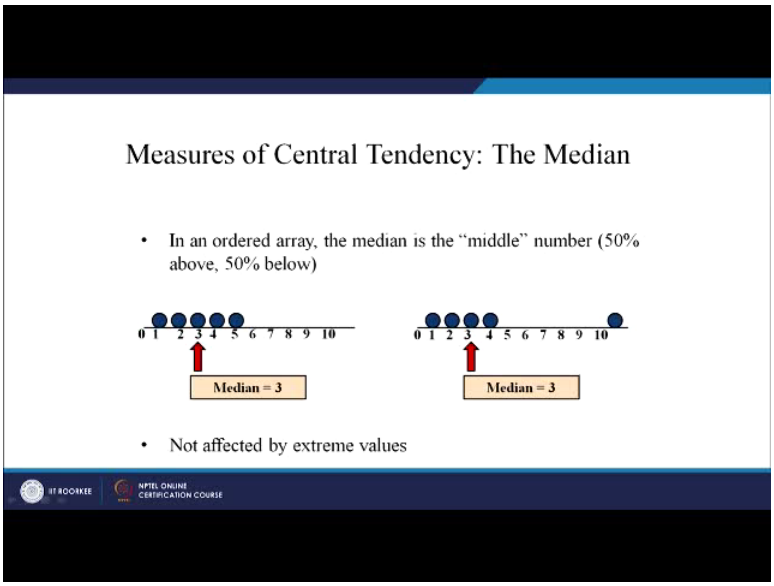
rests, right. Now we have seen mean. Now mean is not a good measure of Central tendency if your data is not normal.

If your data set is normally distributed then mean, mode, median, all would be same, right. But in most of the cases your data would be non normal and then you will not have to use mean. You will have to use some other measure of Central tendency, right. So what is the problem with mean? It is affected by extreme values which we have seen in previous session as well, right. It is tedious to compute lots of data.

So, first of all you take the average and it is a time consuming process. It cannot be computed in case of open class when you got grouped data. Let us say your scale is your class interval is 0-2, 5-10, 10-15 and let us say these are different and let us say 15 to next, 15 onwards This is your open class, all these are closed classes right. This one is open ended class 15 to 15 onwards right. Now it can be 15,000.

So you cannot compute mean if you are mean is falling in this class. So that is why mean is not a, that cannot be calculated for categorical data, isn't it? What is categorical data? That is male and female right you cannot have mean of male and female right. You won't get any answer. So these are couple of disadvantages of mean. So, how to overcome these disadvantages?

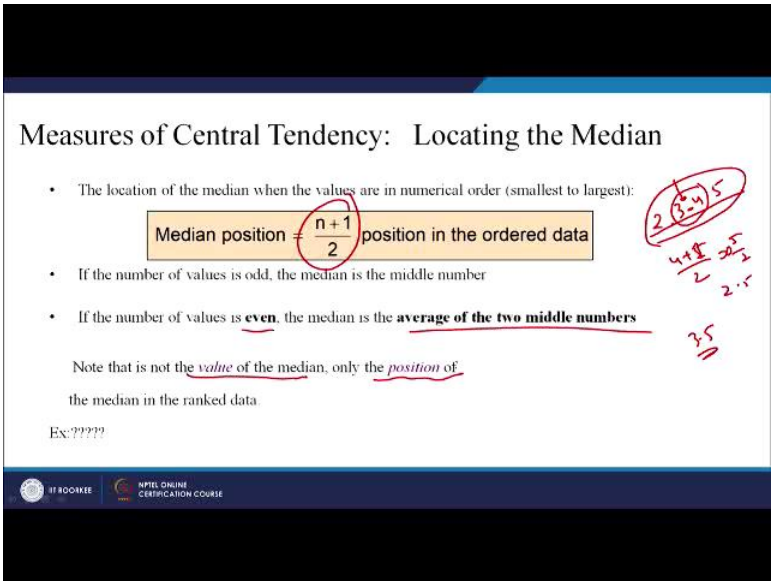
(Refer Slide Time: 13:52)



Let us look at another measure of Central tendency, Median. Median is nothing but the middle point of the data set. So let us say if there are 5 data. Then you should arrange them in ascending or descending order in case of 5 data points, the third one would be the median right. So let's look at these data points on 1,2,3,4 and 5. so median is this data point 3. Median is 3.

Now let us look at another example so, 1, 2, 3, 4 and 10. You just arrange them and in this case they are already in an array form so the median would be 3. It is not affected by Extreme value. So this is the plus point of median right. Now so far as, if you have odd number of data points, it is easy to calculate median.

(Refer Slide Time: 15:01)

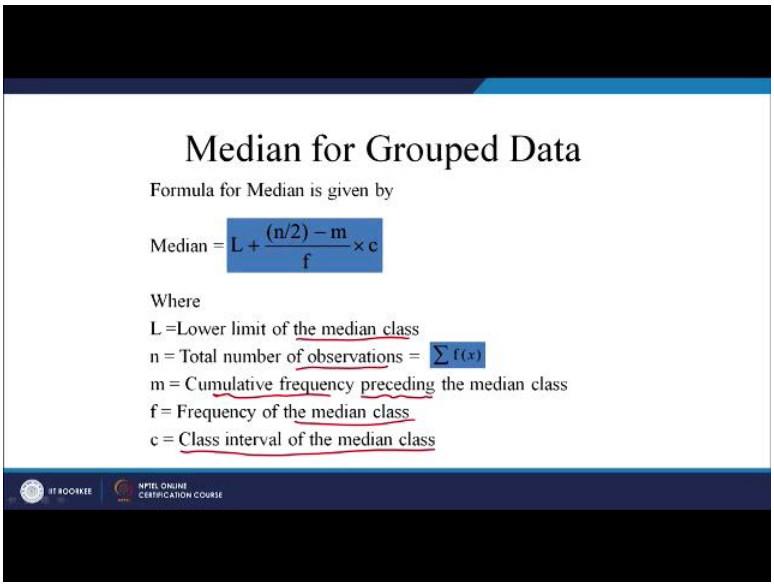


But in case of even number of data points, let us say 2, 3, 4, 5 How would you calculate median, in that case you take this particular formula and it will give you the position of the median, median right. n is equal to 4 in this case right. So 4 + 1= 5 / 2 it is 2.5. It means median would be somewhere here, right somewhere here. But we don't have that data so you just take the average of these two.

So that would be 3.5 median of this particular data set would be 3.5. So this n + 1 / 2 gives you the position of the median not the answer to the question right. If the if the number of values even the median is average of two middle numbers. Note that the value of the median right, it is

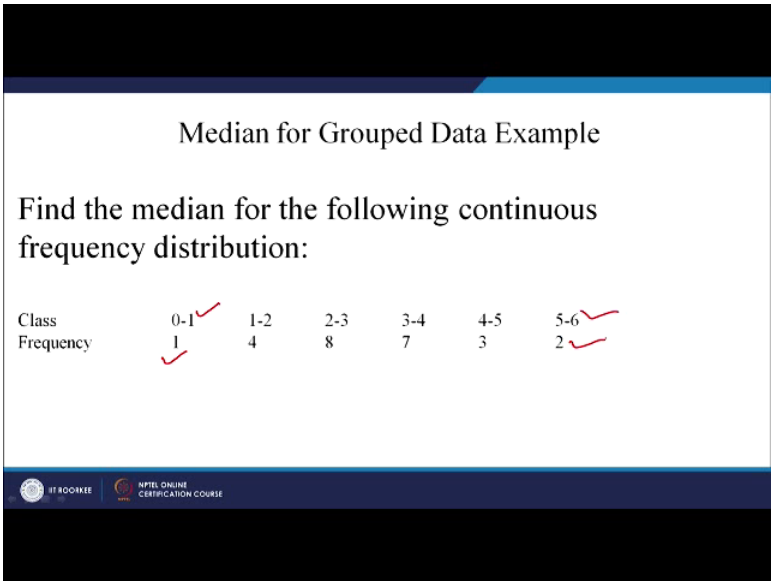
the only position of the median not the answer are not the rank right. So I have already given you example.

(Refer Slide Time: 16:14)



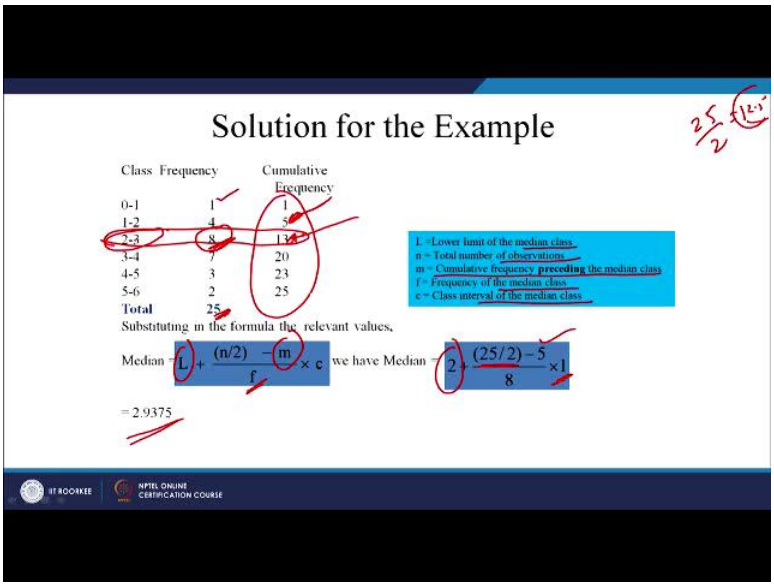
Now how to calculate Median for grouped data. The pervious case was that of ungrouped data. Median is equal to lower limit of the median class and what is Median? Median is 50th percentile value of the data, right, the middle value of the data. In case of even numbers in case of in case of odd numbers but in case of Even Numbers what is the median n+1/2, right. The total number of observations equal to n, m is cumulative frequency of proceeding class, f, frequency of the median class, c class interval. So now let us take an example and calculate median.

(Refer Slide Time: 17:07)



So there are 6 classes to 0-1, 5-6 and you have also been given frequency is of each of these classes right. So this is 1 and this is 2. Now we have to calculate median right.

(Refer Slide Time: 17:25)

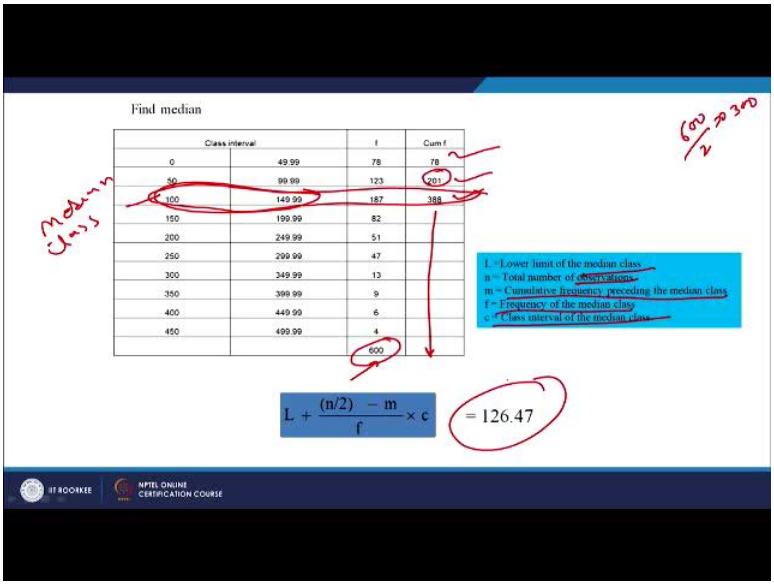


Solution to this example is like this. So, you have got frequencies of each of these classes total is 25. Now what is the median value? It is 50% of the data right. So it is 25 / 2 = 12.5. So where this 12.5 is lying, in case of, in this, this particular column? 12.5 lies here. So this is your median class 2-3 right. Now if you look at this formula L is lower limit of the median class so that would be 2.

Total numbers of observations are 25. So n is 25. 25 by 2 is this right 12.5. -5 what is m? m is cumulative frequency preceding the median class. This is your median class this is your median class and cumulative frequency preceding this is 5. 8 is what? It is f. f is frequency of the median class which is this? Keep in mind that two frequencies one is the cumulative frequency of the preceding median class and this is frequency of the median class and c is class interval of the median class which is one.

Class interval is this 3 minus 2 which is 1. If you put all this values in this formula you will get 2.93. So the median is 2.93.

(Refer Slide Time: 19:15)



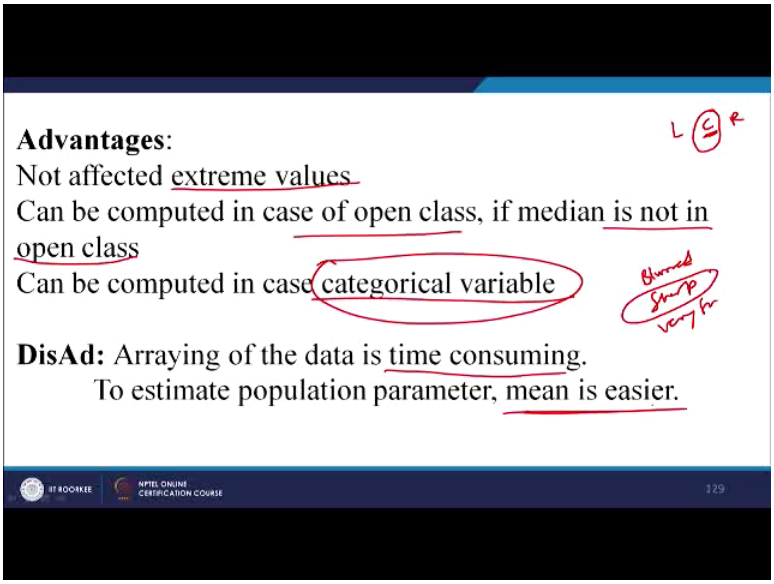
Let us take another example and we will find out median. So this is the same example wherein we have collected data from 600 account holders in a bank and we have seen their balance, right. And this is what we have seen in in previous session as well wherein we calculated mean. So there were approximately 0 to 50 accounts where the balance was 78. The balance was 450 to 500 and there were only 4 such account holders.

Similarly there were 78 account holders who were having balance between 0 to 50. So how would you calculate median. For calculating median what you need you need to first of all find out cumulative frequency? Cumulative frequency is 78 then 201 and 388. You can calculate

remaining cumulative frequencies. So now this total values 600. And what would be the median class or median value would be 300, right.

300 is falling in this particular class right. This is the class. So this is called median class right median class. What is the lower limit of the median class? 100. Total number of observations of course is 600, cumulative frequency of preceding medium class. This is 201 so m is 201. Frequency of the median classes 187 and class interval is 50. So this is how you can calculate answer to this particular question and you should be 126.47.

(Refer Slide Time: 21:25)

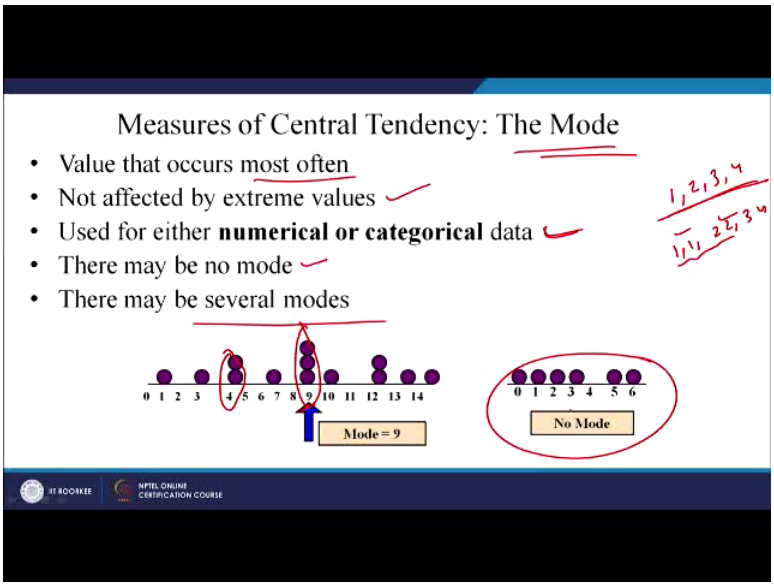


Advantages of median: Of course, not affected by extreme values, can be computed in case of open class, if median is not in that particular open class right; can be computed in case of categorical variables. Now this is important advantage. You cannot calculate mean of categorical variable. But here you can calculate median. To give an example, let us say to give an example let's say there is a blood which is blood, blood sharp and very sharp, right. So this is your medium right.

There are three people sitting so from this is a fellow sitting in left side this fellow sitting right side and this one in centre of this. So this would be the median only the disadvantages you need to arrange data and which is quite a time consuming process. To estimate population

characteristics from sample many a times you will find that mean is easier than median right. Sorry mean is easier than median. So let us move on to next slide.

(Refer Slide Time: 22:54)

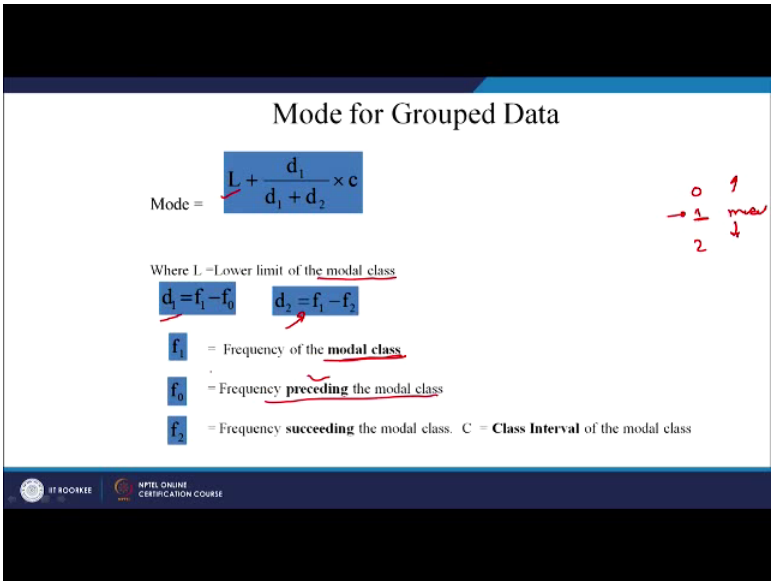


Wherein we look at the third and final measure of Central tendency called mode. Mode is nothing but the value that occurs most often, not affected by Extreme value just similar to what we have seen in case of median used for either numerical or categorical data so this again big advantage there may be no Mode. Yes it is a possibility letter word 1, 2, 3, 4 data 4 data points' right.

So there is no repetition of any data points. So there won't be any mode. And you may have a situation of multimode 1 1 2 2 3 4 so there are 2 modes 1 and 2 modes right. So you can have several modes right. Let us look at this particular example. So 1, 3, 4 its appearing 2 times, 9 appearing three times and so on. So what is the mode? Mode is 9 because that data point is appearing for 3 times.

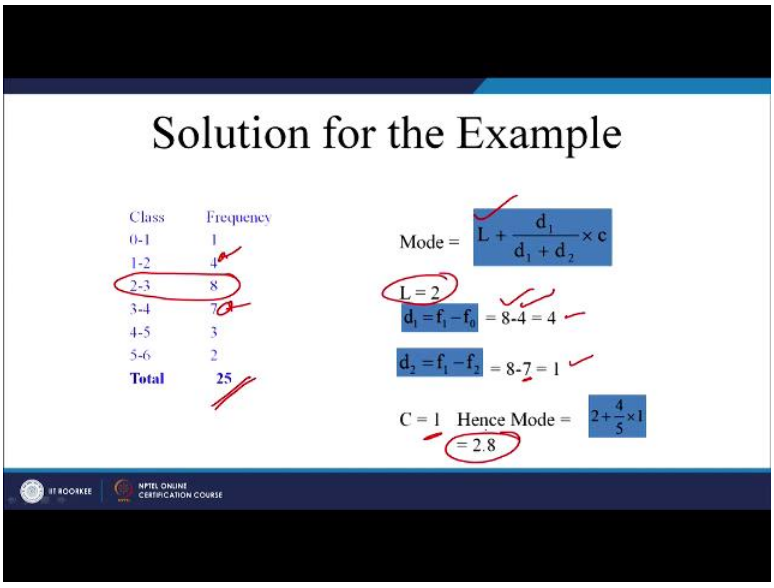
Similarly in this case there is no mode right. No data point is getting repeated. So this is how you can calculate mode in case of ungrouped data. So let us look at mode for grouped data. And again let me remind grouped data means data which are binded by or the frequencies which are having different classes, right. So your grouping data by different class intervals and frequency. That is grouped data.

(Refer Slide Time: 24:46)



The mode is equal to L this is a lower limit of the modal class then you get frequency of the modal class f1, f0 is preceding of the frequency of proceeding model class. So this is very simple, f1 means frequency of the modal class 0 means frequency of previous right 2 means the succeeding class. So this how you can remember. So you need to calculate d1 first, then d2, d1 is frequency of the modal class minus frequency of the preceding model class. Similarly d2 is f1minus f2.

(Refer Slide Time: 25: 38)



So, let us look at this question which is having different classes and frequencies. So how would you solve this question? So this is 25 right; sum of the frequencies. Now the modal class is what

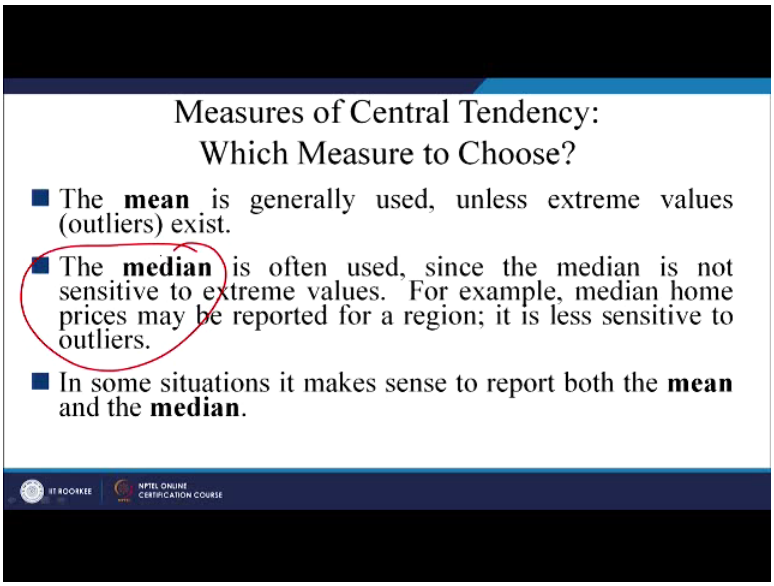
is mode wherein you have got frequency. This is most right. So this is your model class. So class interval is 1 right. So L is lower limit of modal class which is 2 right; f1, what is f1 if one is look at this previous slide. f1 is frequency of the modal class.

So frequency of model class is 8. So f1 is 8; f0 is frequency of the price of the previous model class right. This is the previous model class previous to the modal class. Similarly 7 is what? Frequency of the next class after modal class 7. So 8-4 =4, 8-7=1 and C is what? Class in interval right isn't it. Class interval is 1 right. So this is 1. Hence the answer is 2.8.

So you can easily calculate. So let me summarize what we have done in today's session. We have seen how to find out weighted mean, we have seen how to find out geometric mean, we have seen how to find out median of grouped and ungrouped Data, Mode of grouped and ungrouped data. So at the end of the day which one will you select? Mean, mode or median does that is the decision to be taken.

As I said if your data are normally distributed then all will be same. But in real life situation you will have most of the times non-normal data. And in case of no normal data you can say you will always have skewed data with positively or negatively skewed, so you should select median, ok as the measure of central tendency, ok.

(Refer Slide Time: 28:07)



So with this let me complete today's session. We will have few more sessions in which we will look at the measures of dispersion thank you very much