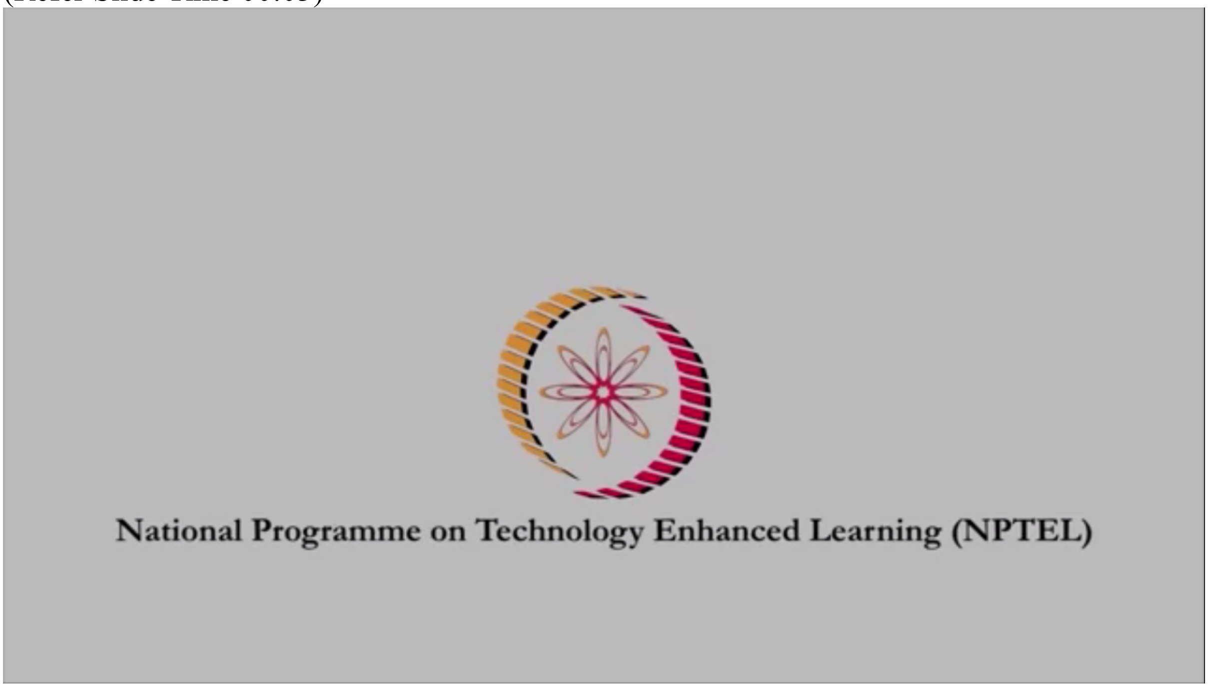
(Refer Slide Time 00:01)

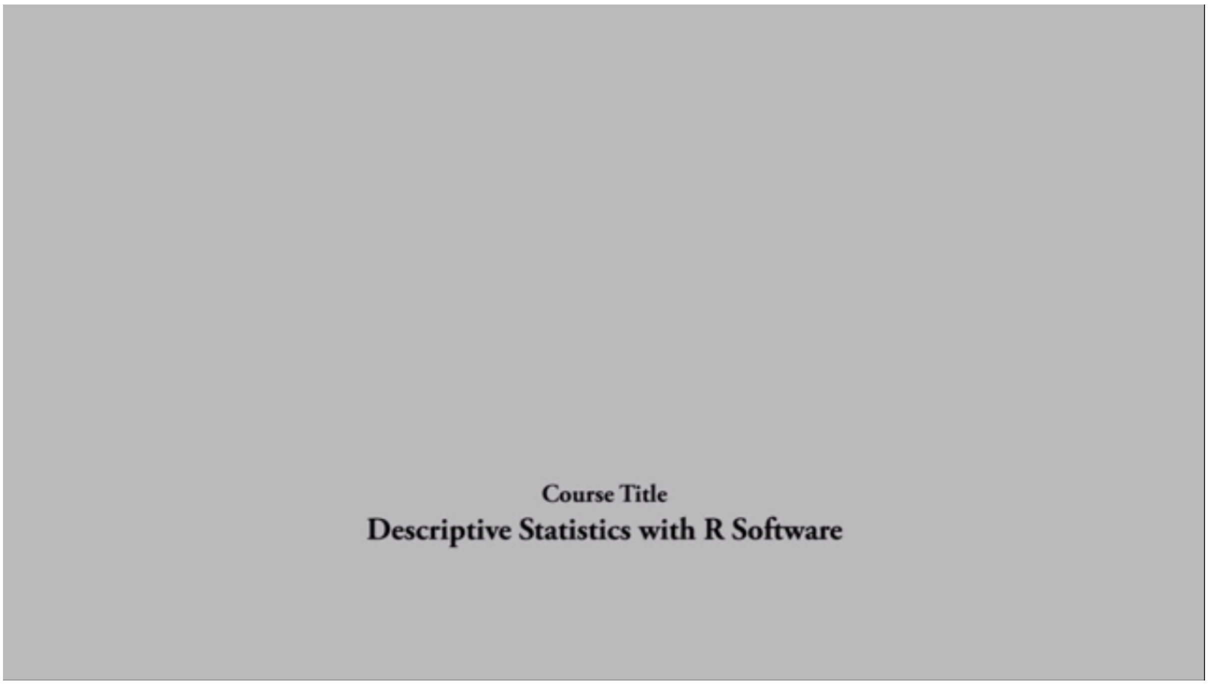
Indian Institute of Technology Kanpur

(Refer Slide Time 00:03)



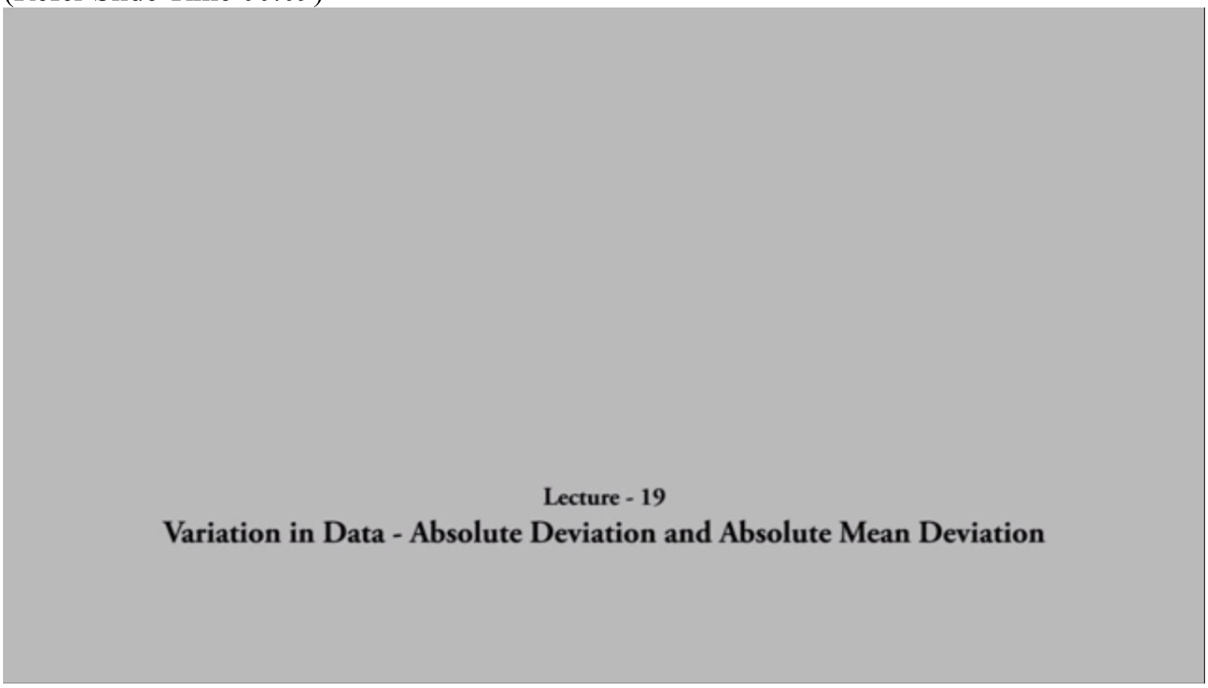
National Programme on Technology Enhanced Learning (NPTEL)

(Refer Slide Time 00:06)



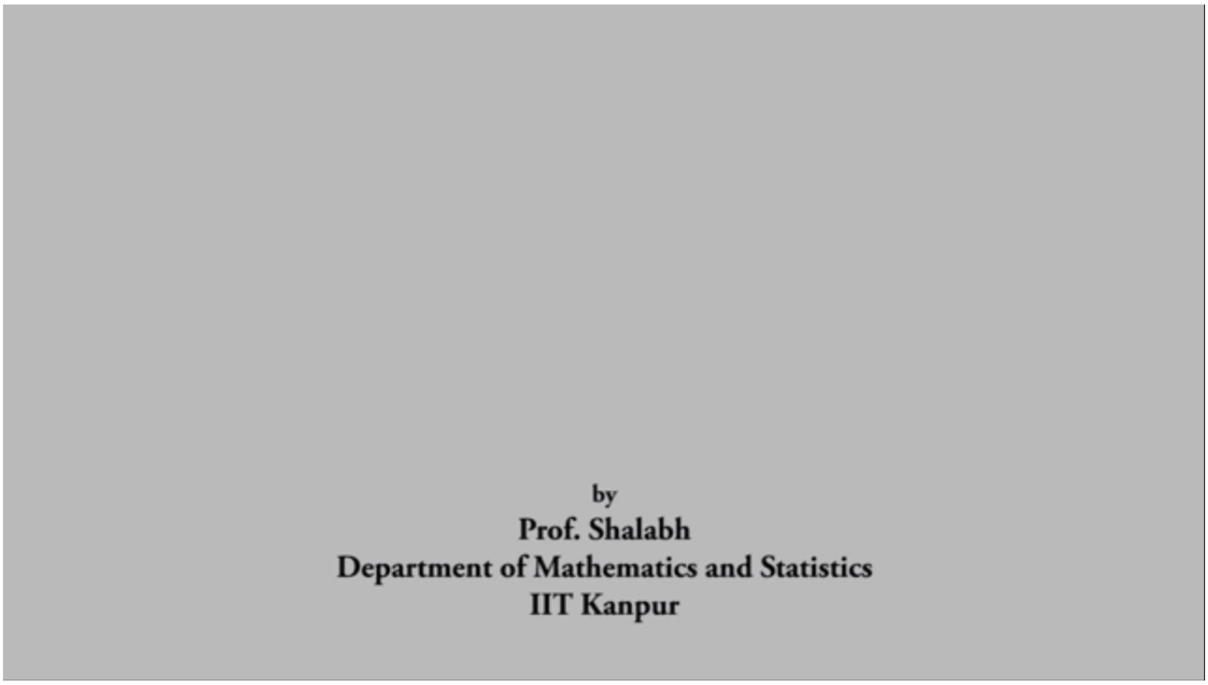
Course Title
Descriptive Statistics with R Software

(Refer Slide Time 00:09)



Lecture - 19
Variation in Data - Absolute Deviation and Absolute Mean Deviation

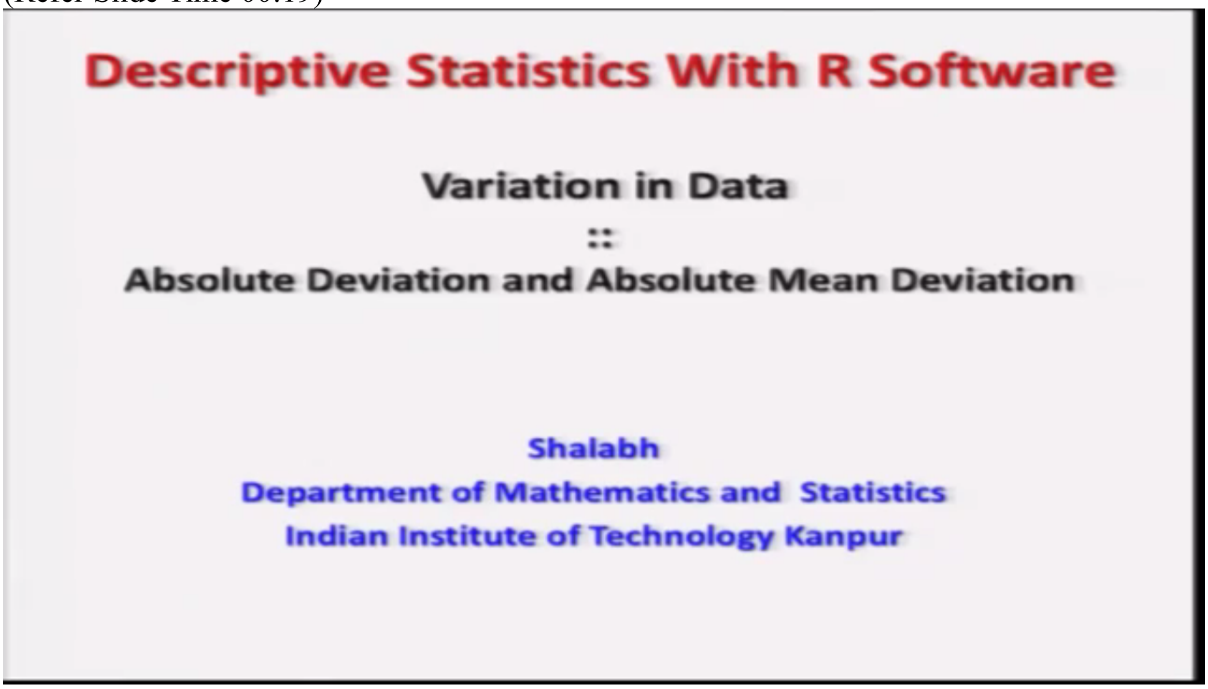
(Refer Slide Time 00:11)



by
Prof. Shalabh
Department of Mathematics and Statistics
IIT Kanpur

Welcome to the next lecture on the course Descriptive Statistics with R Software.

(Refer Slide Time 00:19)



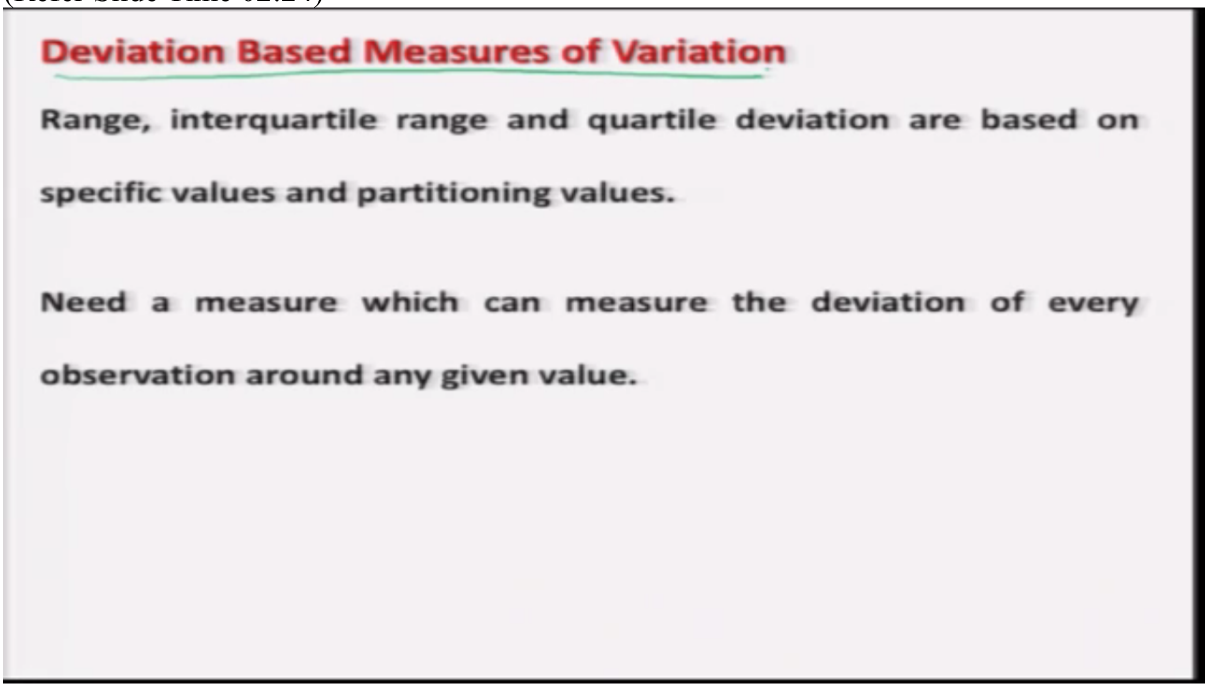
You may recall that in the earlier lecture, we started our discussion on the topic that how to measure the variation in the data and in that lecture we had considered three possible tools: range, quartile deviation and interquartile range. The measure of range was dependent only on two values, that is the minimum and maximum, and quartile deviation as well as interquartile range. They were dependent on two values: the first quartile value and third quartile value.

So now there is another thing that these measures are going to be based only on say two values at a time, either minimum maximum or first and third quartiles. They are trying to take care of the entire dataset in a different way. That is the minimum maximum quartiles are computed on the basis of all the observations.

But now there is another concept to measure the variation that why not to measure the variation of individual data points from the central value and then try to combine all the things together; combine all the information, combine all the deviation together.

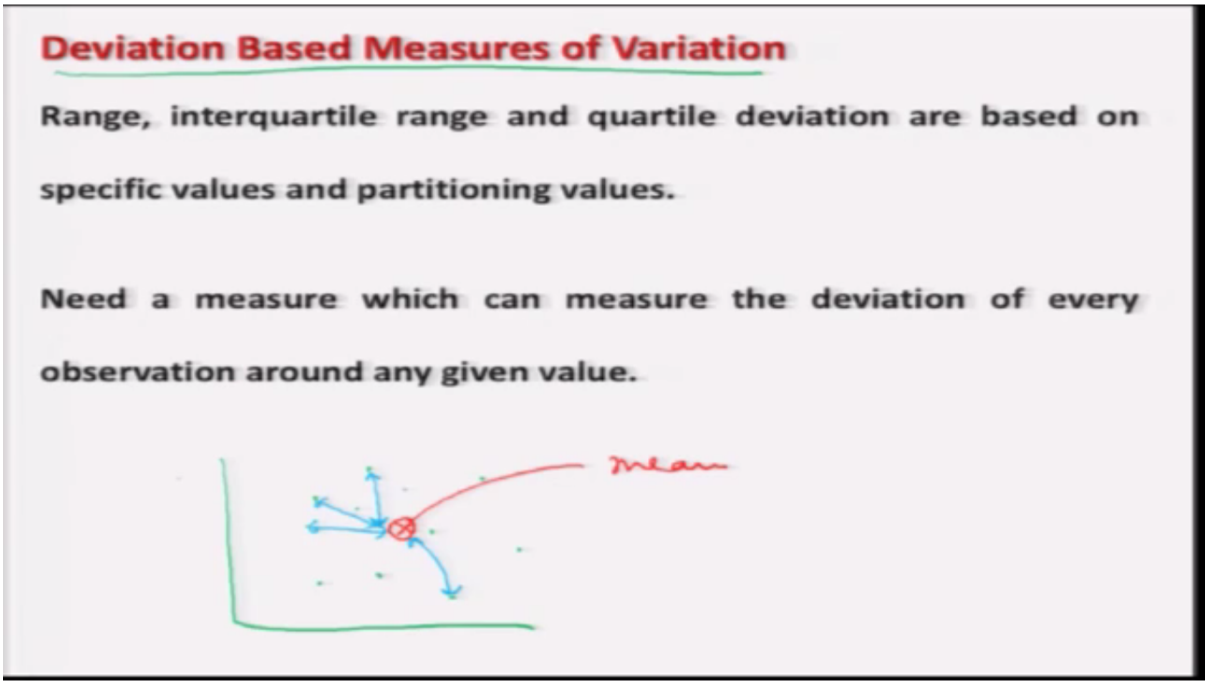
So now we are going to start a discussion on different types of tools to measure the variation which are based on the individual deviation of the data from the central value or any other value also. So, in this lecture, we are going to discuss on two topics. First is absolute deviation and another is absolute mean deviation. We will try to understand the concept and I will try to show you how to compute it on the R Software also.

(Refer Slide Time 02:24)



Now we are going to discuss the another aspect that what are the different measures which are based on the deviation? What does this mean? If you remember in the early lecture, I have made this type of data, right, and if you try to see here, the mean value is somewhere. I'm assuming somewhere here. This is the mean value and I'm trying to make this difference, this. So these are my data points and this blue colour lines are indicating the difference between the central value, say mean or it can be another value also from this individual data points. So these are the deviations.

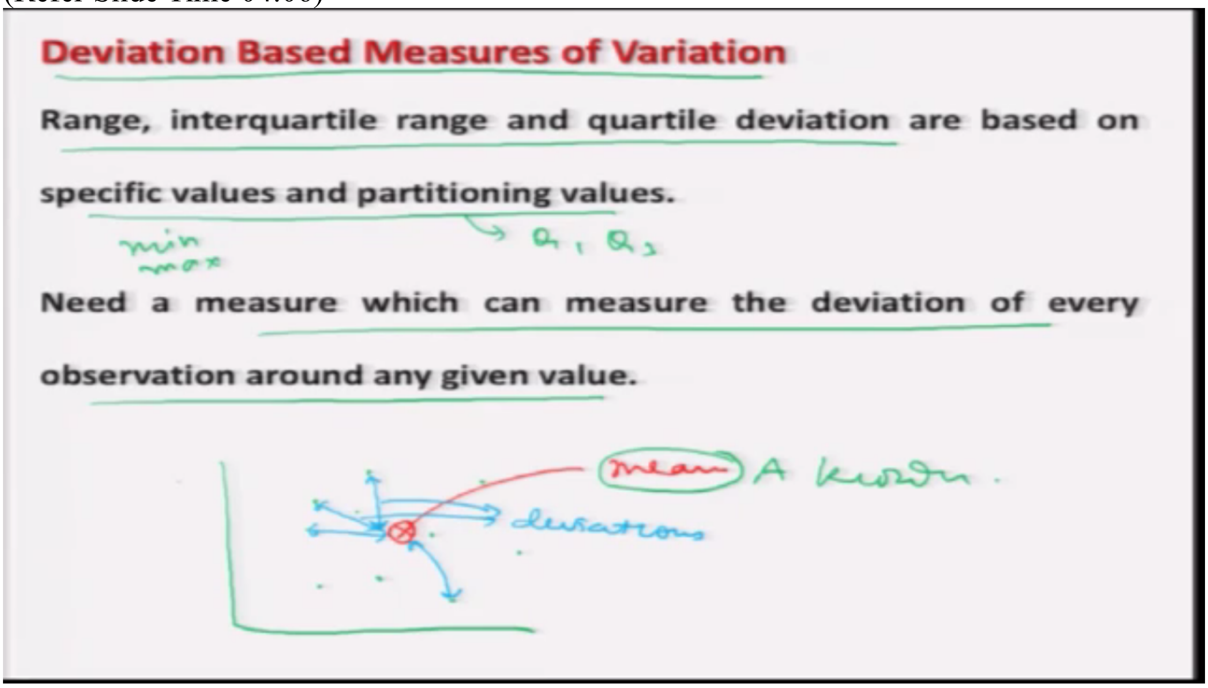
(Refer Slide Time 03:16)



So now the concept what we are going to discuss here is that the tools like range, interquartile range, and quartile deviations, they were based on specific values and partitioning values. A specific value mean minimum value, maximum value where the partitioning values were first quartile, third quartile.

Now we would like to have a measure which can measure the deviation of every observation around any given value. So in this figure instead of taking here mean, I can also take it here say any known value say here A. So how to get it done?

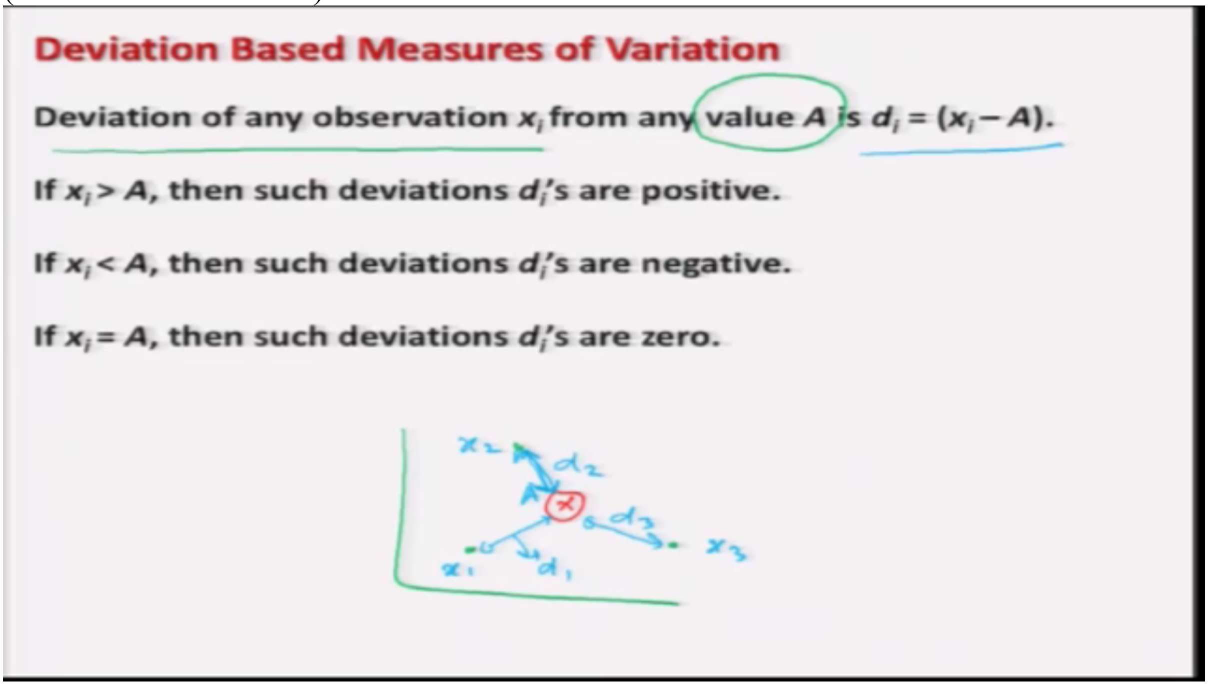
(Refer Slide Time 04:06)



So now if you say the first question is how to measure the deviation, and first I am initiating the discussion that I would like to measure the deviation around any value A and later on I will try to choose appropriate value of A. So you can see here that in this graph, if my this is the data point and this is here the mean value, then or the known value here A, then these are my deviation.

So this is my here x_1, this is x_2, and this is x_3. And essentially, this difference is trying to measure the difference between x_1 and say here let me denote it by here d_1. And similarly, this difference here, this is trying to measure the deviation between x_2 and A which I am trying to denote it here as d_2, and similarly, the different between here this A and x_3 denoted by here d_3. So, in general, I will say here that $d_i = x_i - A$ is going to measure the deviation.

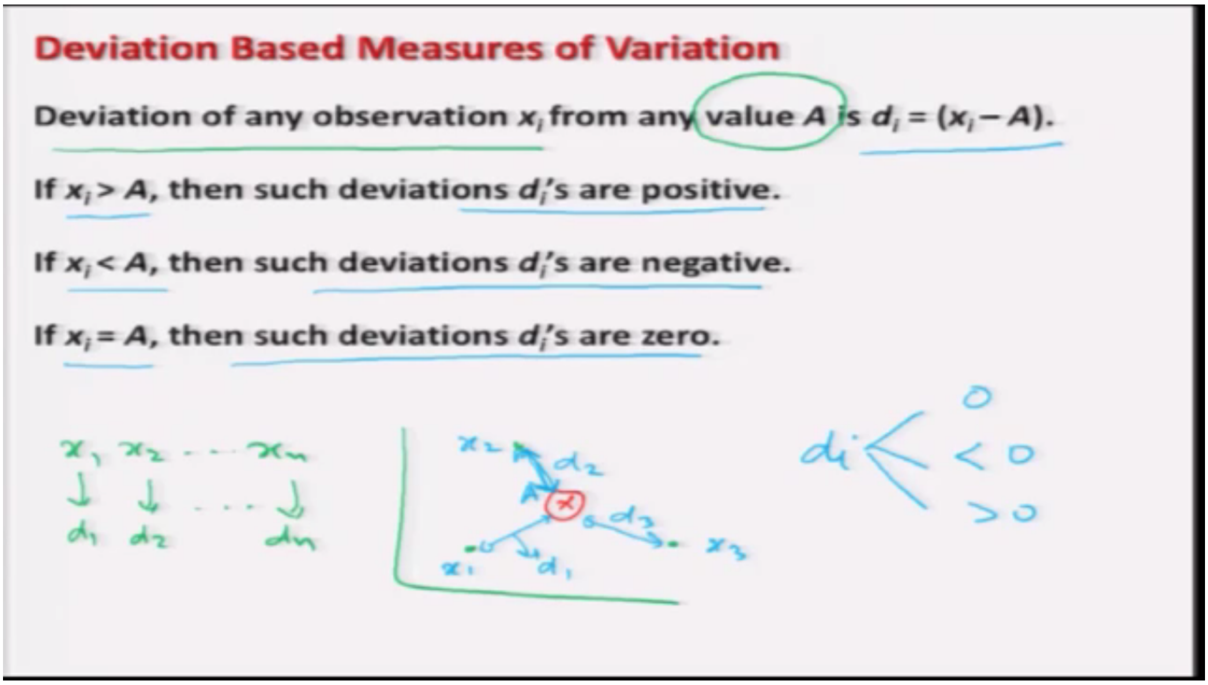
(Refer Slide Time 05:23)



Now in case if x_i is greater than A, then all such observation where this whole, in those cases the deviations will be positive, and in case if the value of x_i is smaller than A, then all such deviations will be negative, and if x_i is exactly equal to A, then all such deviations are going to be zero. So now this d_i can take three possible values: zero, less than zero, and greater than zero.

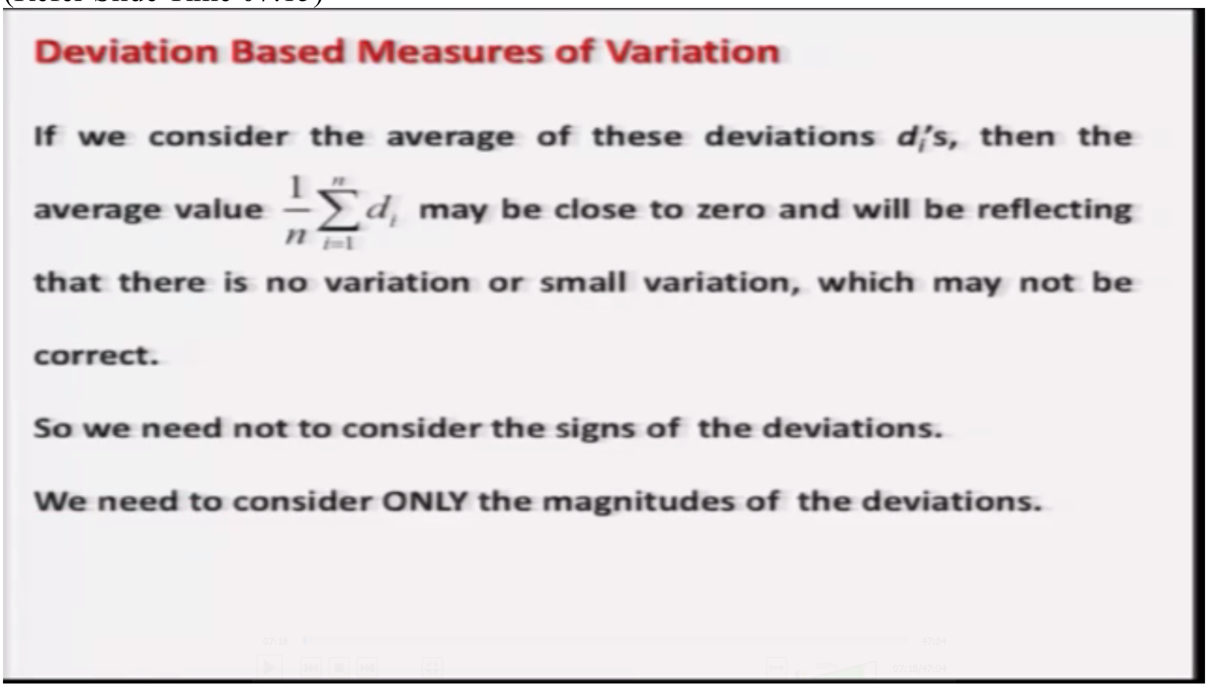
Now there is another issue. The issue is this that in case if I have got the observations say x_1, x_2, x_n, then corresponding to every observation I have got the value of deviation d_1, d_2, d_n.

(Refer Slide Time 06:15)



Now d_1, d_2, d_n, they are the individual values. So suppose you have got 20 observations and then you will have 20 values of deviation. Some are positive. Some are negative. Some are zero. But it will be very difficult by looking at deviation and get a summarized value. As we had discussed in the case of measures of central tendency that the basic tendency is that we would always like to have a summarized measure. That means all the information which is contained in d_1, d_2, d_n, this has to be somehow combined in a single quantity by looking at which I can always compare the things. Right?

(Refer Slide Time 07:15)

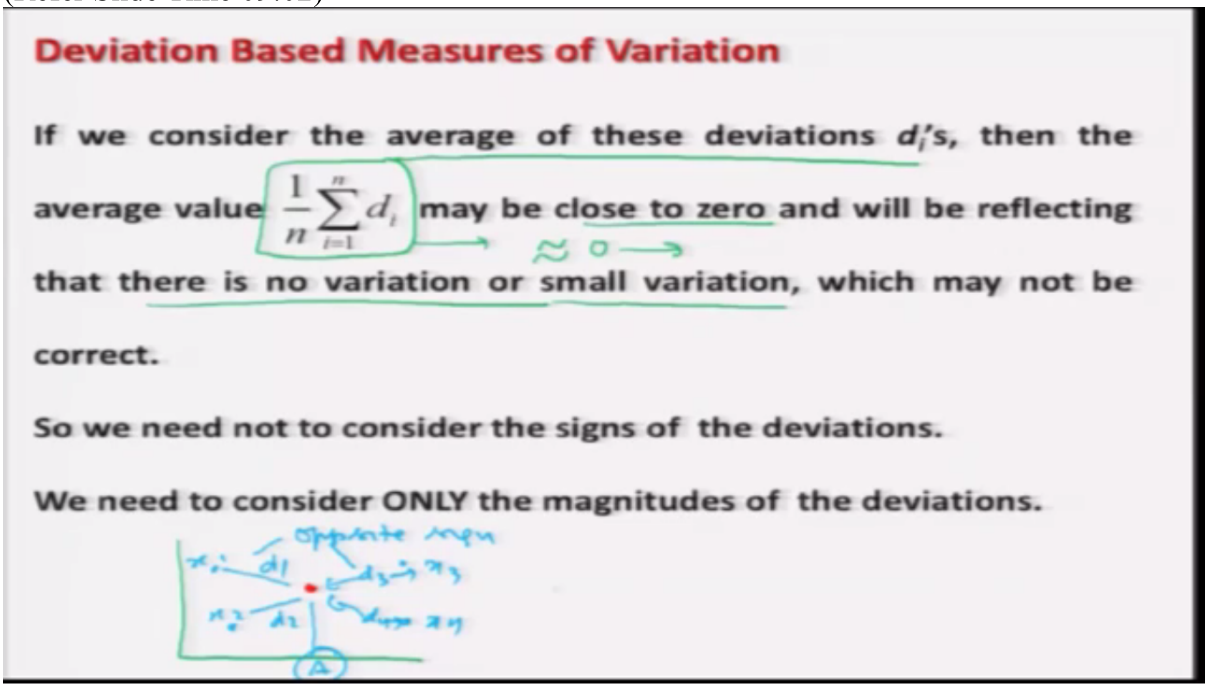


So now one option is this that once I have the values of d_1, d_2, d_n, I can find out the average of these deviations. So I try to compute summation of d_i divided by n, but this deviation may be very, very close to 0. Why? Because it is possible that some of the deviations are positive;

some are negative. So there is a possibility that when we try to take the average of positive and negative values, the mean may come out to be 0 or very, very close to 0. So once this value is coming out to be zero that summation d_i upon and this is 0, then it is possibly indicating as if there is no variation in the data or if this value is pretty small that means it is indicating that there is very small variation in the data.

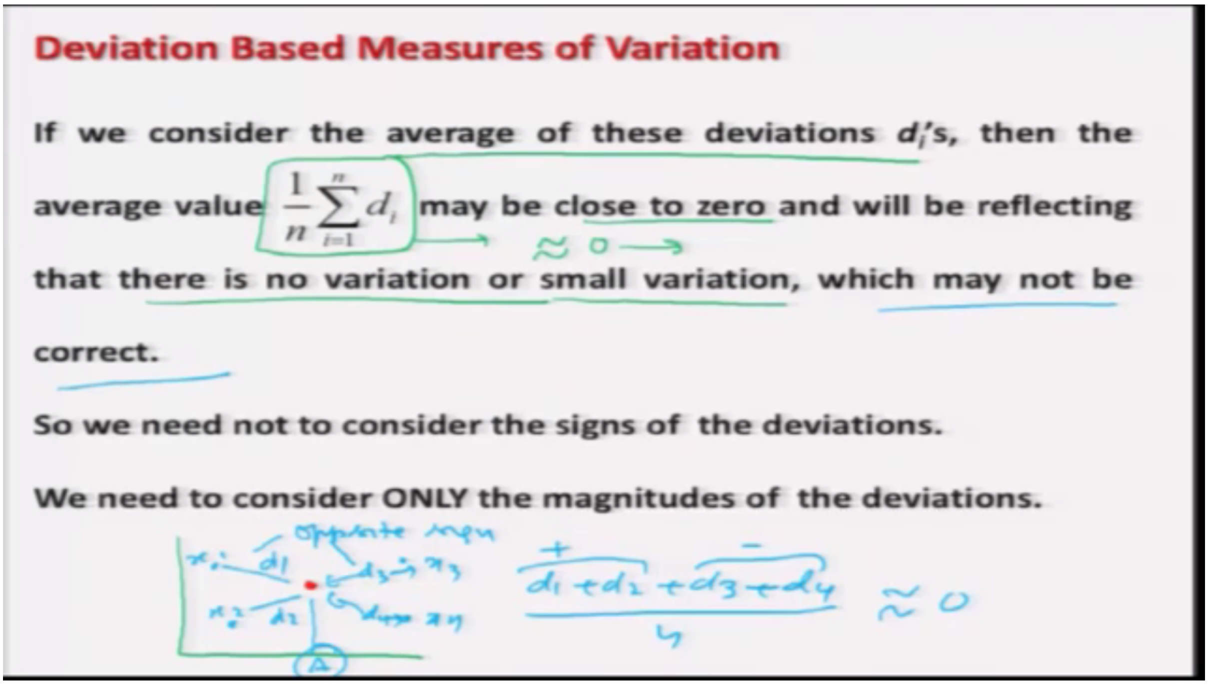
So if I try to make here a figure here that suppose my central value is somewhere here and my observations are something like as one and two are here and one and two are here. So in this case if you try to see here if this is my here A and x_1, x_2, x_3, x_4, then the deviation d_1, d_2, and deviation d_3 and d_4, what will happen? d_1, d_2 will have the opposite sign as of d_3, d_4 considering that d_1, d_2 had been measured with respect to A.

(Refer Slide Time 09:02)



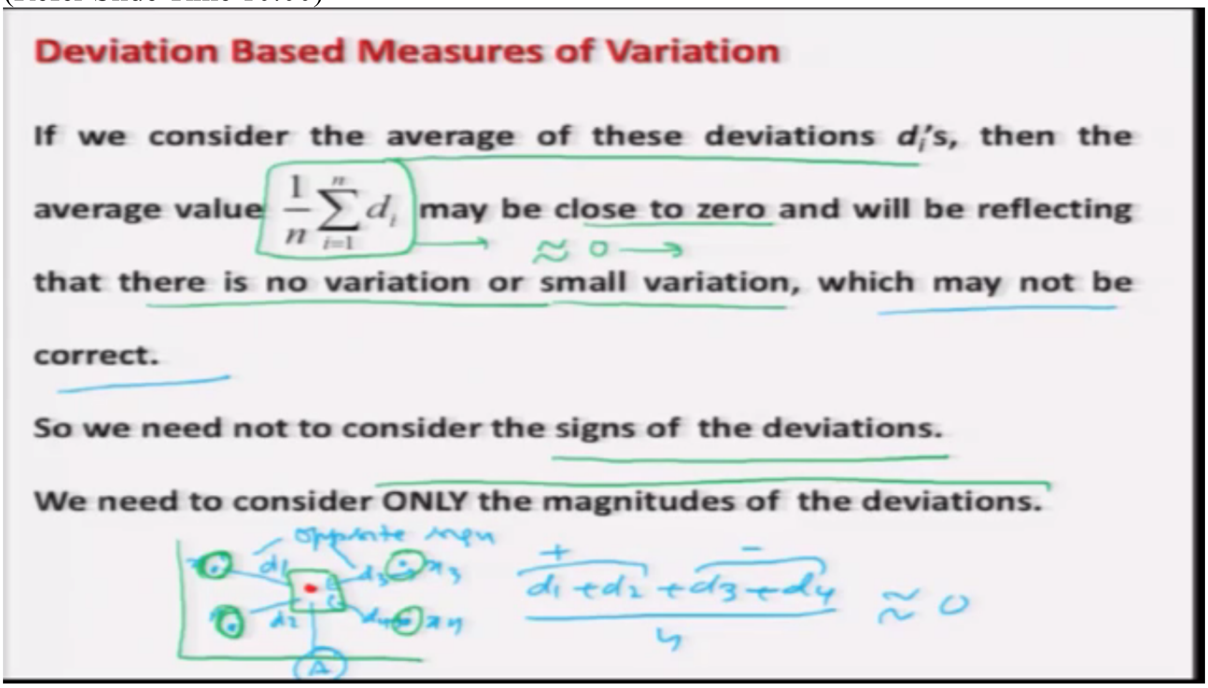
So now in case if I try to find out the value $d_1 + d_2 + d_3 + d_4$ divided by 4, then two are suppose positive and two are suppose negative values, then their average may be close to 0 or exactly 0. So this may be misleading because may not give us the correct information.

(Refer Slide Time 09:24)



So, obviously, now by this example, it gives us a clear idea that we need a measure where these signs are not considered. Why? Because I'm interested only in the scatteredness of these green circles around this red point. I'm not interested in their individual values. So we need not to consider the signs of this d_i, signs of these deviations, but we need to consider only the magnitudes of these deviations.

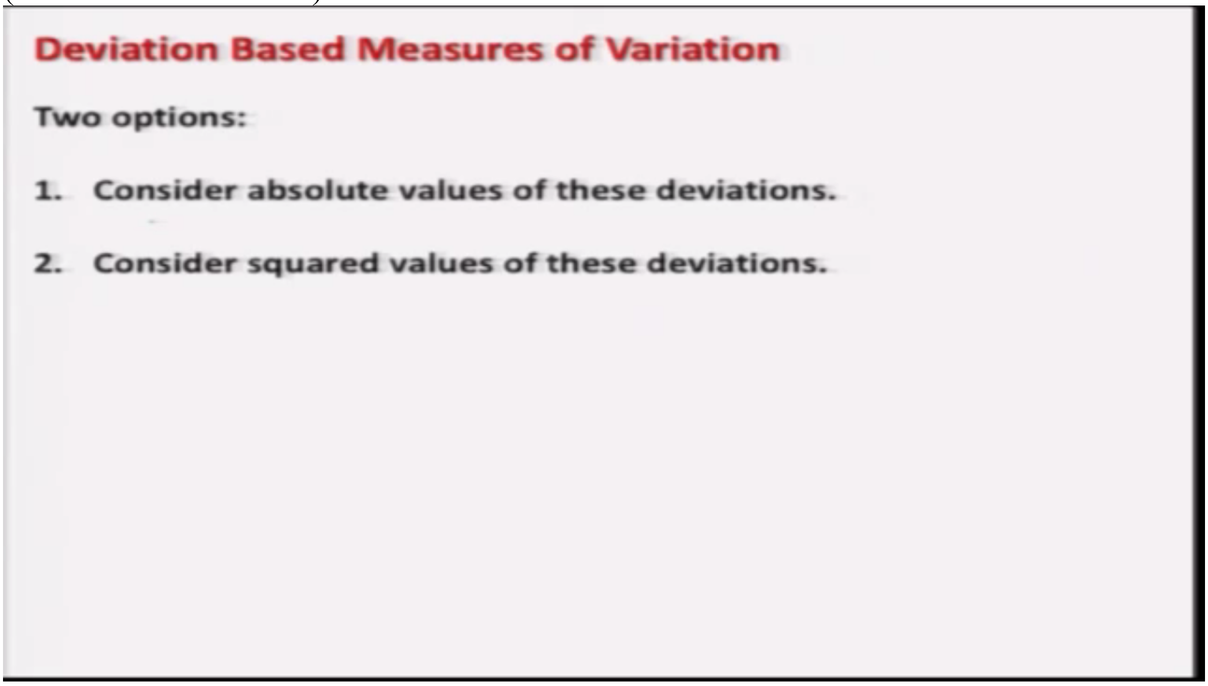
(Refer Slide Time 10:00)



So now after looking at this example, we have understood that now I need a summary measure which can denote the deviation in the data or the variation in the data based on the deviation and these deviations have to be considered only in their magnitude, but obviously, in the data some deviations are going to be positive and some deviations are going to be negative. So as long as we have positive deviations, there is no problem, but then I'd need to convert the negative deviations into a positive deviation. That means I'd need to convert the sign from negative to positive.

Now the next question is how to convert the sign? So in mathematics we have understood that there are two options. Either I try to consider the absolute value or I try to make the negative values to be squared. So now based on these two aspects, we have two types of measures. One is absolute deviation and another is variance. So, in this lecture, I'm going to consider the concept of absolute deviations and in the next lecture I will try to consider the concept of variance.

(Refer Slide Time 11:25)



So I will be considering two options to make the negative deviations to be positive or to consider their magnitude. Consider absolute values or consider their squared values of those deviations.

So now in this lecture I'm going to concentrate on absolute values. So I had the observations x_1, x_2, say here x_n. Now I have taken the deviation from the value A. Now after this I will try to take the absolute value of all these deviations and now I need to combine all this information into a single measure.

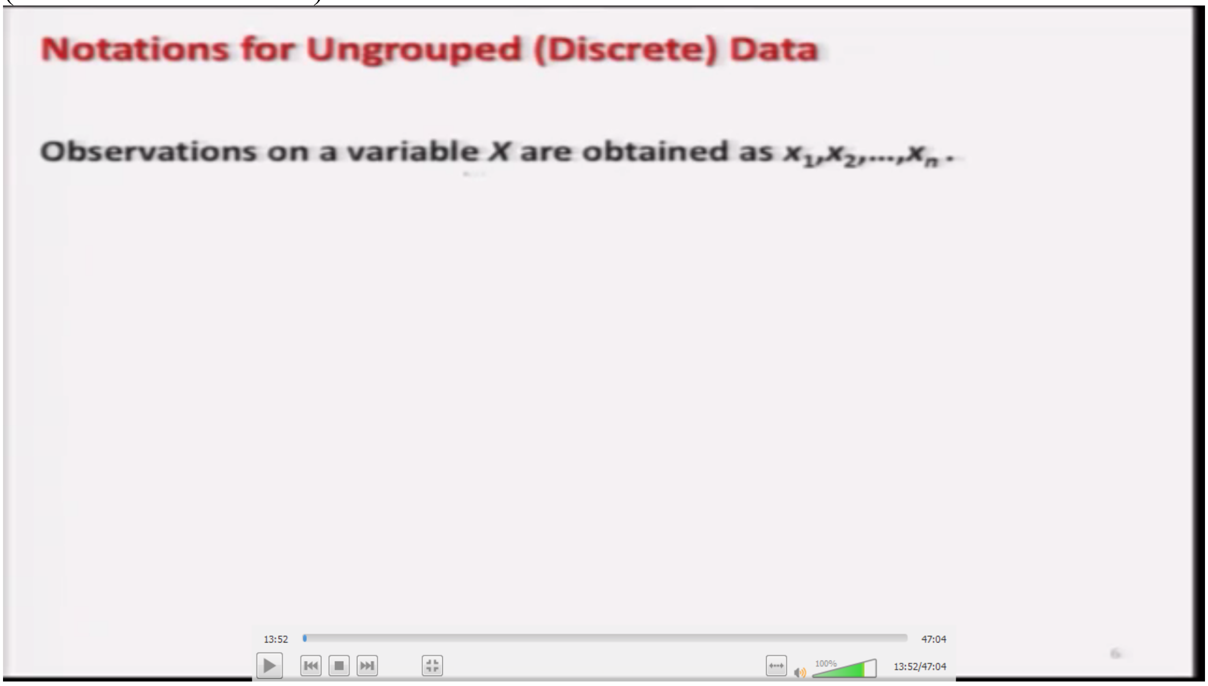
Now the question is this, how to combine it? So we will consider it. Now before going to the discussion on the how to combine such information, first, let me clarify the symbols and notation what I'm going to use in this lecture and the next lecture. You see I have two types of data sets. One is on continuous variable and another is on discreet variable.

In discrete variable, we simply try to use the observations as such, but in the case of continuous variable, we try to group them. We try to convert the data into a frequency table and then we try to extract the mid-values of the class intervals and the corresponding frequencies to construct our statistical measures, the same thing that we had done in the case of arithmetic mean if you remember.

For the ungrouped data that was defined as summation of x_i divided by number of observations and in the case of continuous data or the group data, it was defined as summation x_i f_i divided by summation of f_i and in the second case x_i's were denoting the mid-values of the class interval.

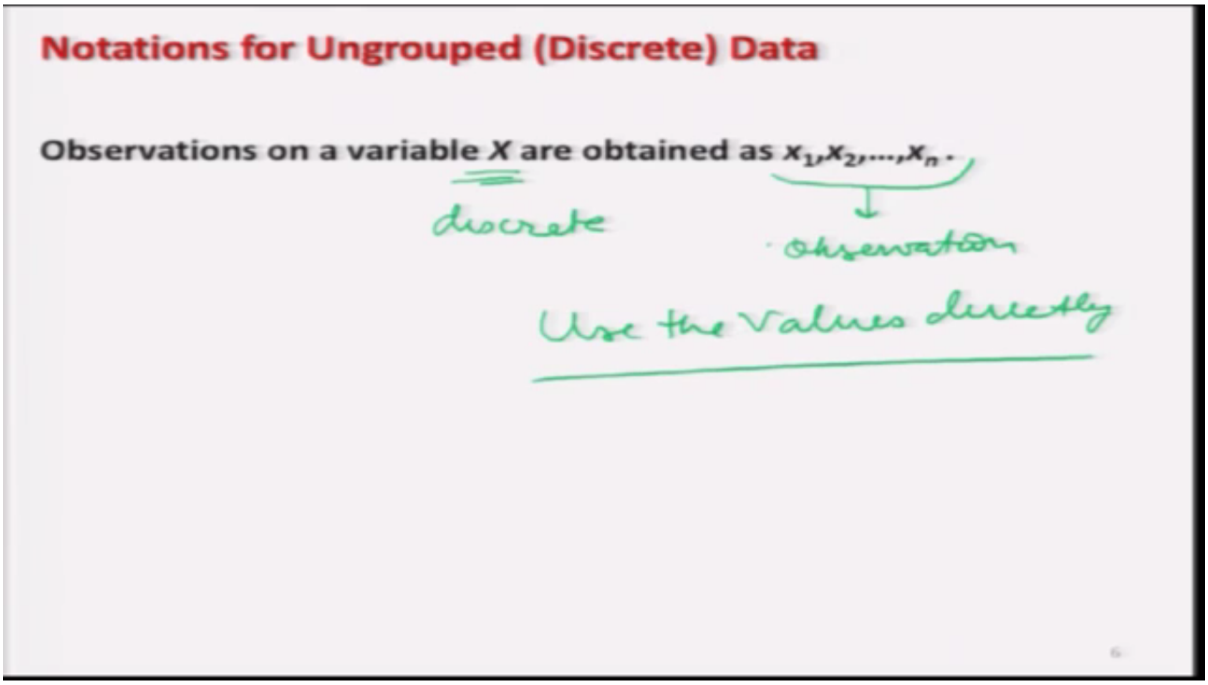
So I would now try to introduce these measures for grouped data as well as for ungrouped data.

(Refer Slide Time 13:51)



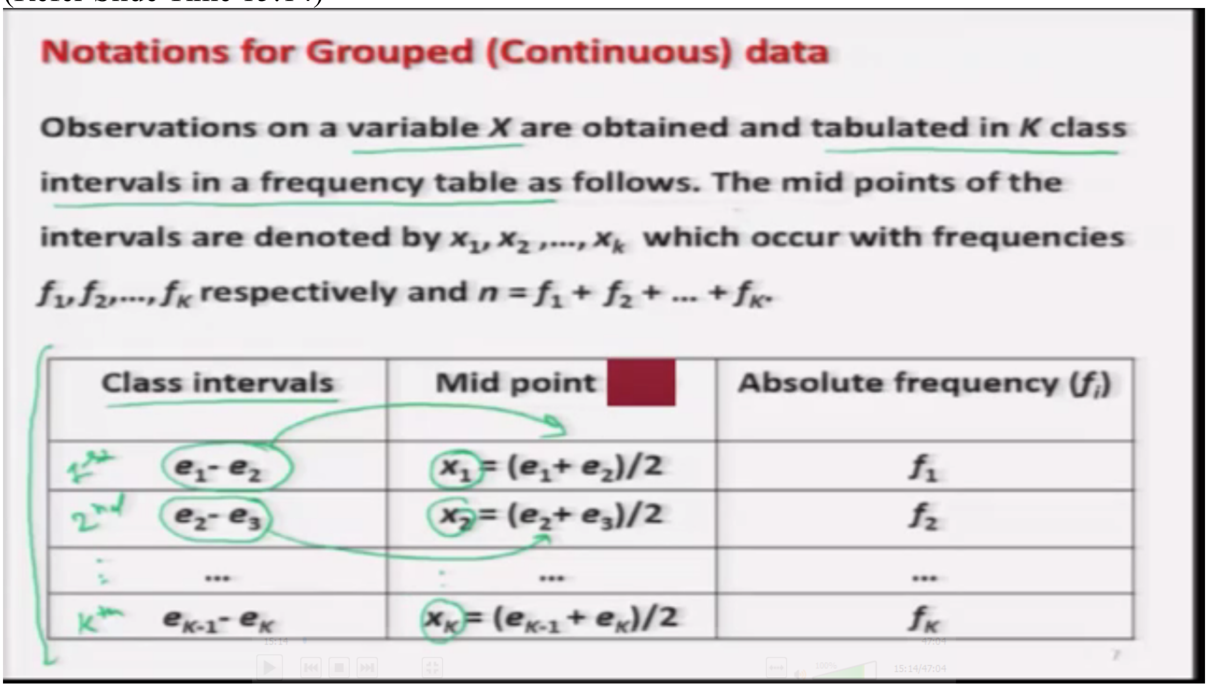
So, first, let me just explain here that in case if our variable is discrete or the data is ungrouped, then we will convert the sets of set of observation as say x_1, x_2,...,xn. There is no issue and we will use these values directly, use the values directly.

(Refer Slide Time 14:17)



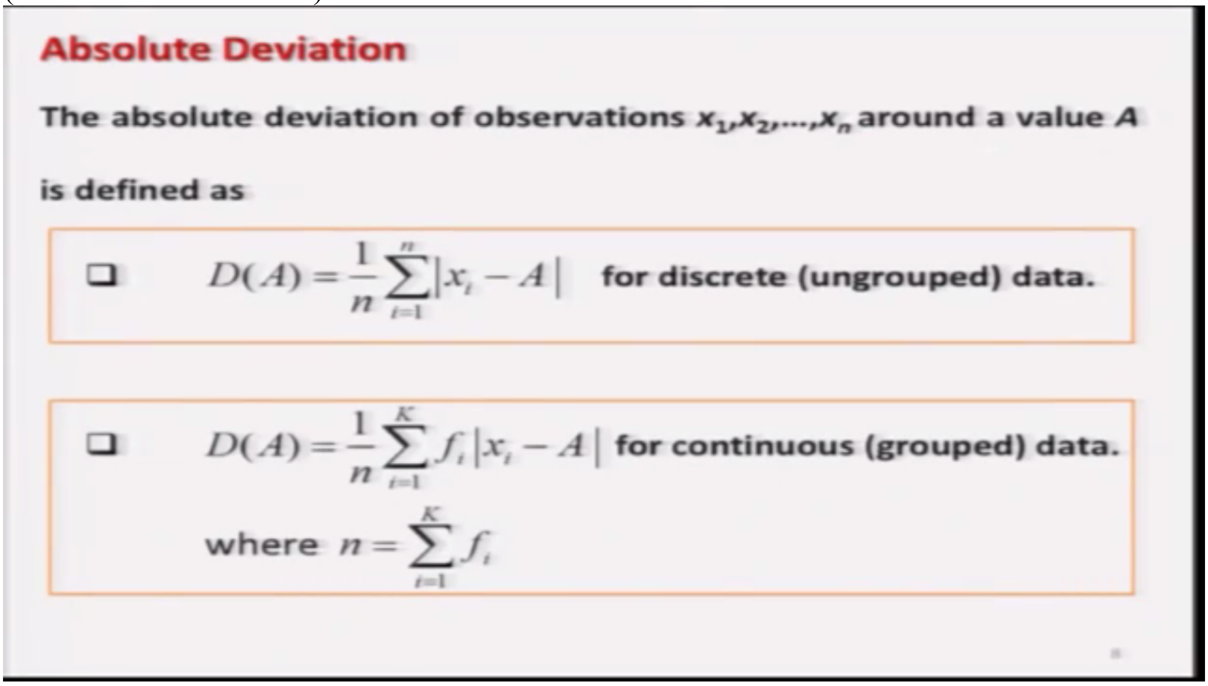
Whereas in case if I have grouped data, so suppose my variable here is X and I have got the observations, which I have tabulated in K class intervals in the form of a frequency table like this here. If you can see you here, I have made the observation in this class interval, say, first class interval, second class interval, up to here, K[th] class interval and after that I have found the mid points of this class intervals. So midpoint of the class interval e_1 to e_2 is rather $(e_1 + e_2)/2$. The class interval e_2 to e_3 has the midpoint at $(e_2 + e_3)/2$ and these mid points are going to be devoted by x_1, x_2 and so on up to here x_K. Right?

(Refer Slide Time 15:14)



So I am converting the mid points of the intervals by $x_1, x_2, \ldots, x_k$ and which is just different than in the earlier case where we had denoted $x_1, x_2, \ldots, x_k$ as the values that are obtained directly from the data and corresponding to these intervals we have the absolute frequencies, which I will be denoting by f_1, f_2, f_K and the sum of these f_1, f_2, f_K is going to be denoted by n, which is again different from the earlier case from the case of ungrouped data.

(Refer Slide Time 16:01)

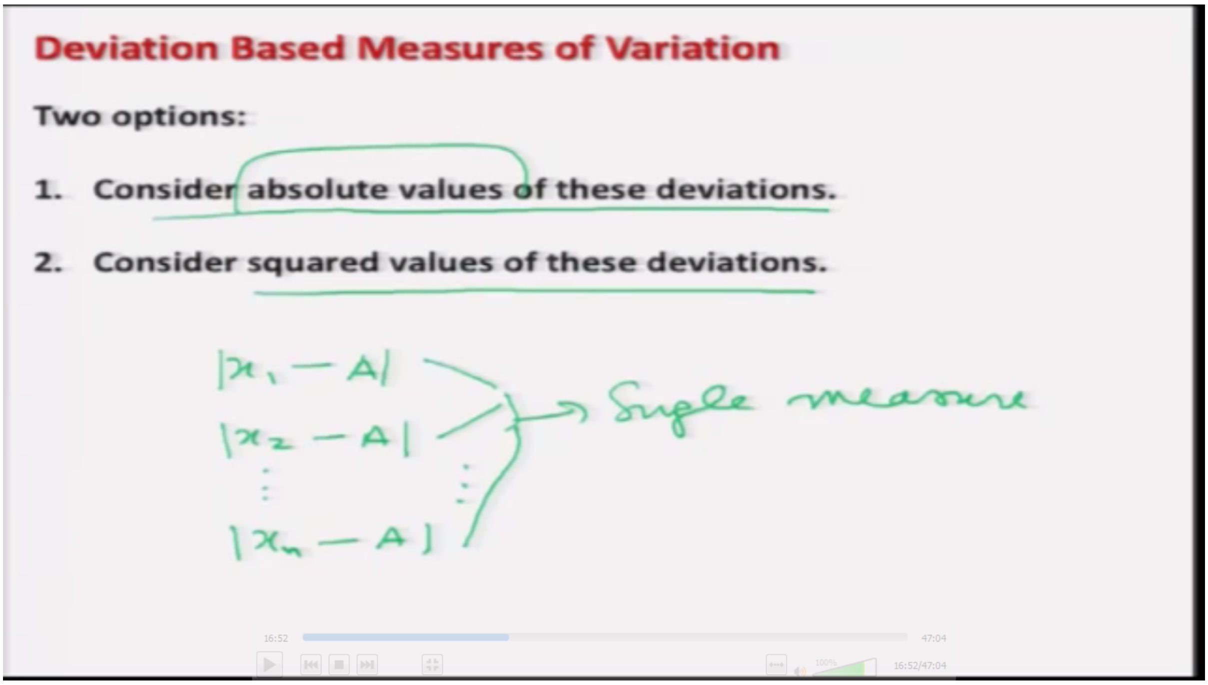


So this is what you have to keep in mind that when I'm trying to consider the case of ungrouped data, then x_1, $x_2, \ldots, x_k$, they are going to denote the values which I'm observing directly, and when I'm trying to consider the ungrouped data, then x_1, $x_2, \ldots, x_k$ are going to denote the mid-values of the class intervals and the corresponding frequencies in those class intervals are f_1, f_2, f_K.

Similarly, the symbol is small l that is going to denote the number of observation in the case of ungrouped data whereas a small n will be denoting the sum of all the frequencies in case of grouped data.

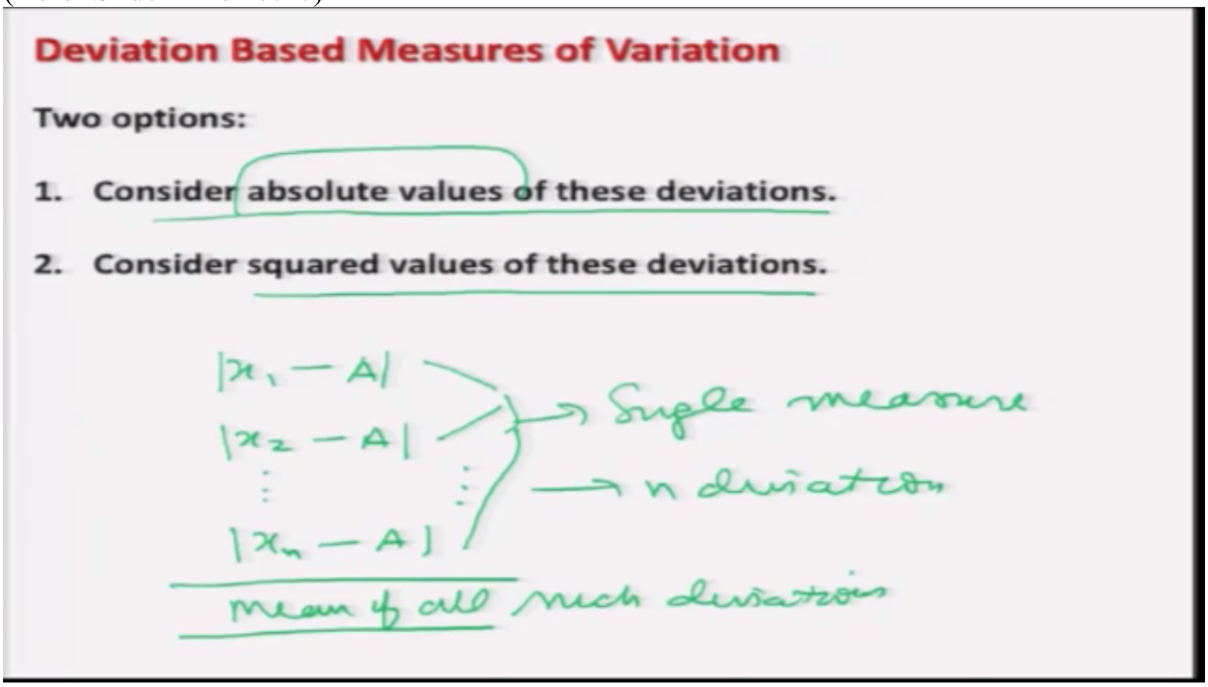
So now with these notations I will start the discussion on the first measure, that is absolute deviation. Right?

(Refer Slide Time 16:51)



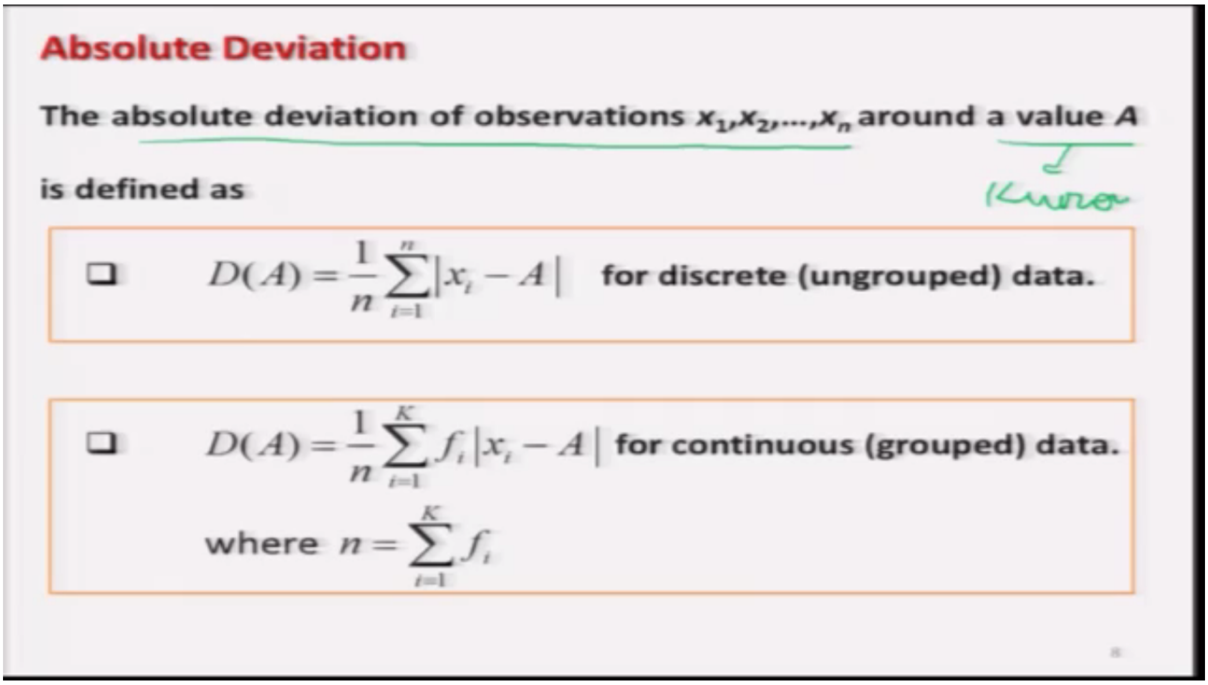
So, first, try to concentrate on this slide, this one here. You can see here I have obtained here this n such deviations which are the absolute values. Now in case if I try to find out the mean of all such deviation, then this will give me a sort of summary measure of absolute deviation.

(Refer Slide Time 17:18)



And the definition of absolute deviations is that if you have got a data x_1, x_2,...,x_n, then the absolute deviation of this observation around a value A, which is known, is defined as follows.

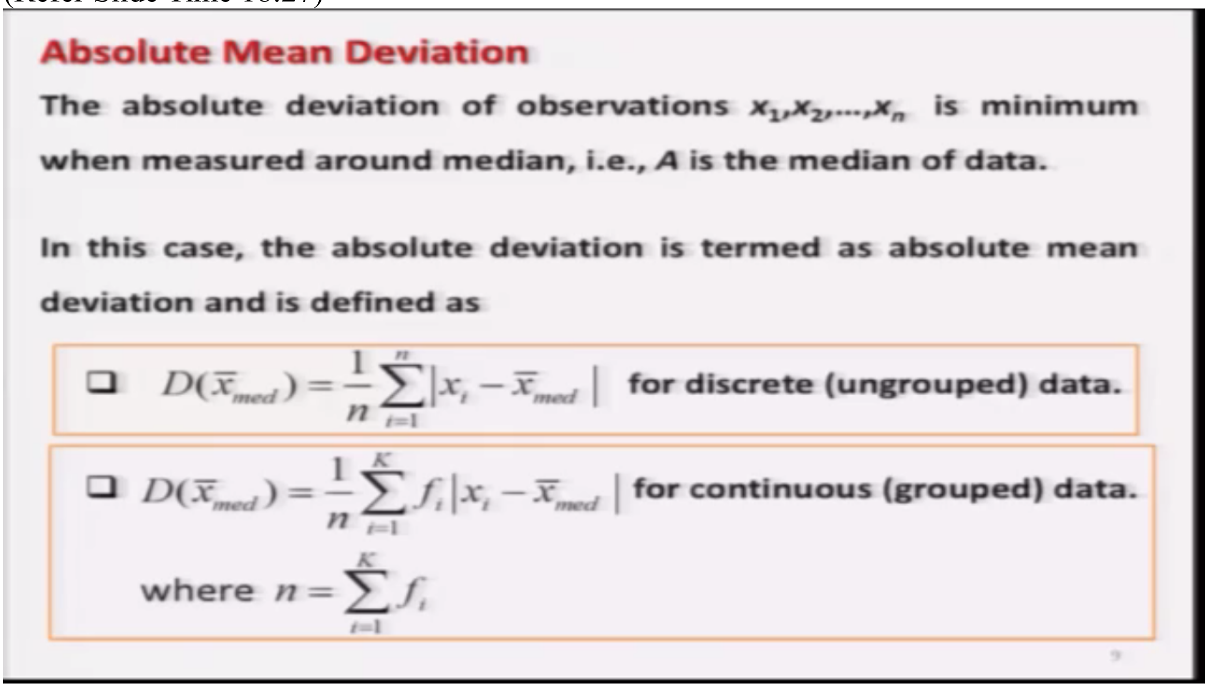
(Refer Slide Time 17:37)



Now I have here two cases for ungrouped or discrete data or for grouped or say continuous data. In case of discrete data, we simply try to take the arithmetic mean of such absolute deviations and in case of continuous data, we again try to find out the mean of such absolute deviation using the concept of arithmetic mean in the case of frequency table.

So now you can see here, this is how we try to define the absolute deviation, sum of absolute deviation divided by n, which is here the arithmetic mean of the absolute deviation and in the case of continuous data also this is again the arithmetic mean of the absolute deviation which are suitably defined. Right?

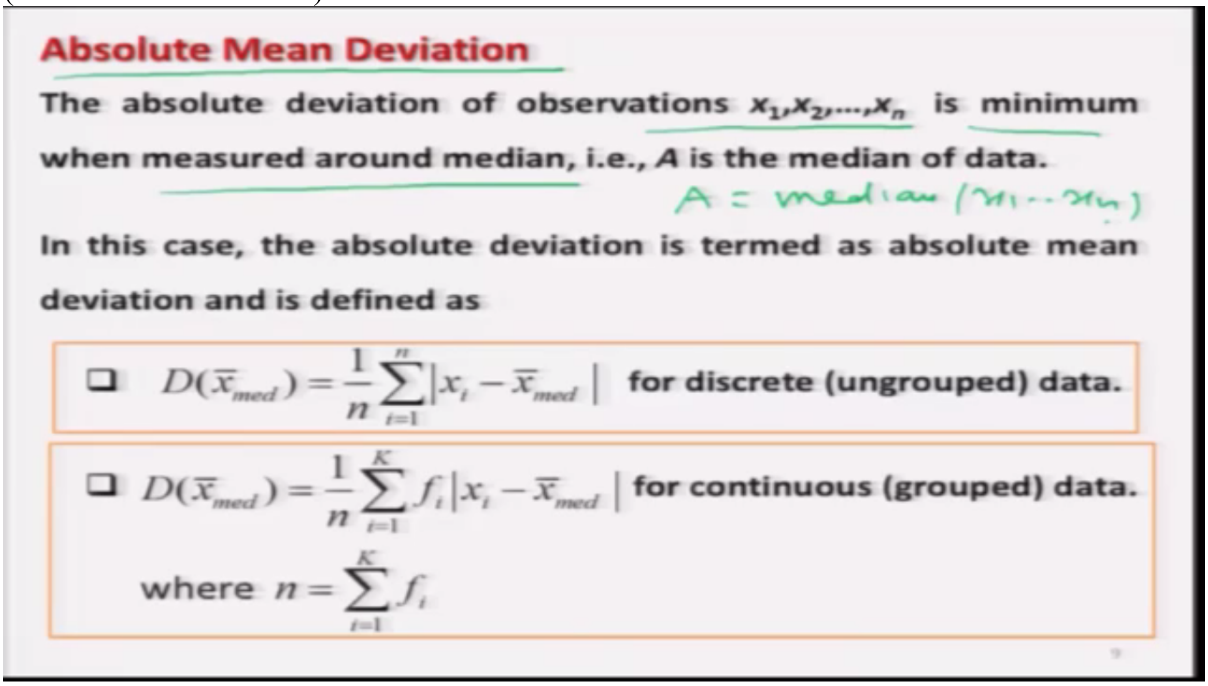
(Refer Slide Time 18:27)



Now the next question comes, how to choose this A? Well, in this definition, I'm assuming that A is known to us, but in case if I try to choose different values of As, then the values of absolute deviations are going to be changed. So by using a small algebra, we find that these absolute deviations is going to take the smallest value when this A is chosen to be the median of those values. So we try to replace the value of A by the median of $x_1, x_2,...,x_n$ and this gives us another measure, which is called as Absolute Mean Deviation.

So in case if I try to take the observations $x_1, x_2,...,x_n$, then this absolute deviation is going to be minimum when it is measured around median. That is A is the median of the data, say, $x_1, x_2,...,x_n$.

(Refer Slide Time 19:36)

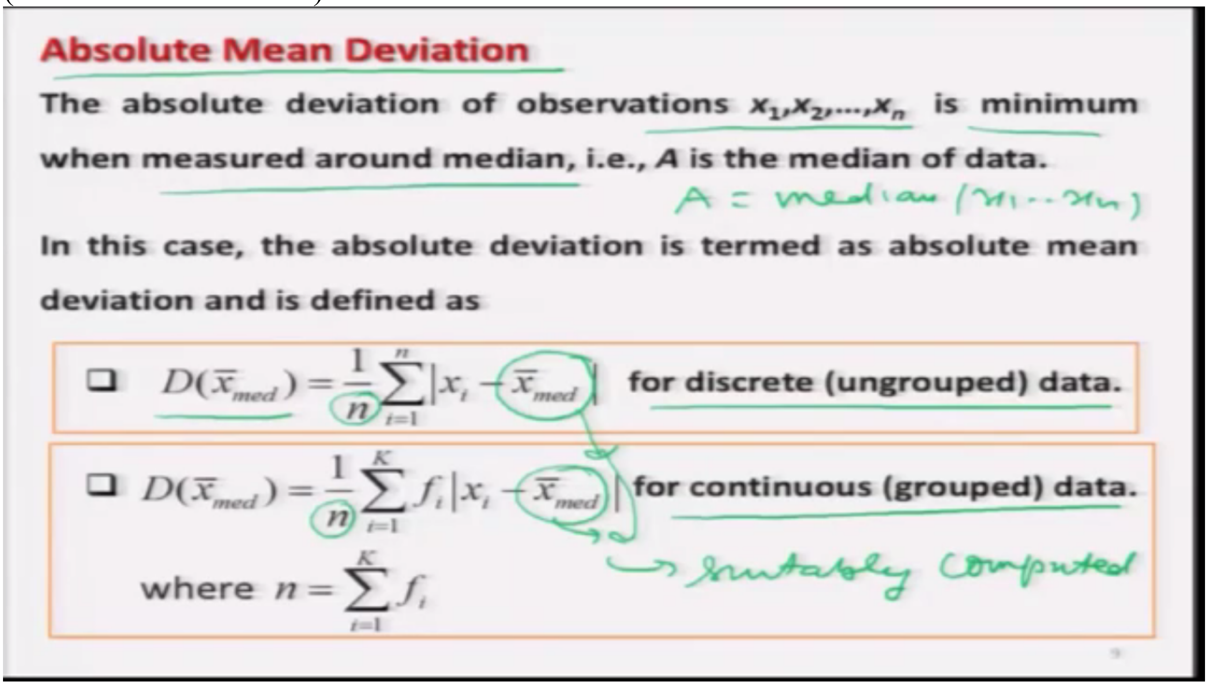


Remember one thing. We had computed the median in case of a grouped data and ungrouped data by different expressions. So whenever we're trying to find out the value of absolute mean deviation, then we need to compute the value of median suitably. Suitably means if you have discrete data, try to order the observation and then try to take the middle value depending on whether your number of observations are odd or even what we have discussed in the case while handling the topic of median.

And similarly, if you have grouped data, then you please try to convert it into a frequency table and use the expression for computing the median from that table, and you remember that we had computed the median in a different way and by not using the direct functions. Right? So this is what you have to keep in mind.

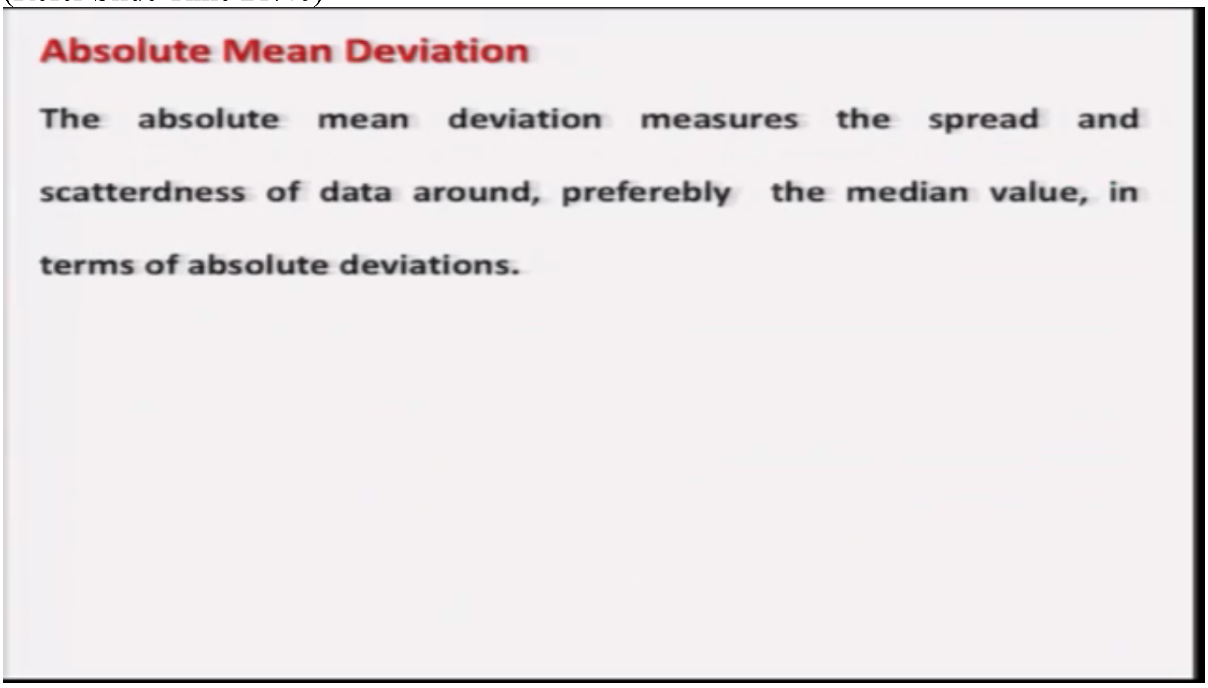
So in case if I have a discrete data, then I simply try to replace A by the median of $x_1, x_2,...,x_n$, which is denoted by $\bar{x}_{med}$ and then I try to take the arithmetic mean of those values and this is called as the absolute mean deviation in case of ungrouped data. And similarly, if you have grouped data on continuous variable, then, again, I try to find out the value of absolute deviation by replacing A equal to here by $\bar{x}_{median}$, but remember this $\bar{x}_{median}$ and $\bar{x}_{median}$, they are suitably computed. Right. And also note that the definition of here n and the n in the case of grouped and ungrouped data, they are going to be different. Right.

(Refer Slide Time 21:43)



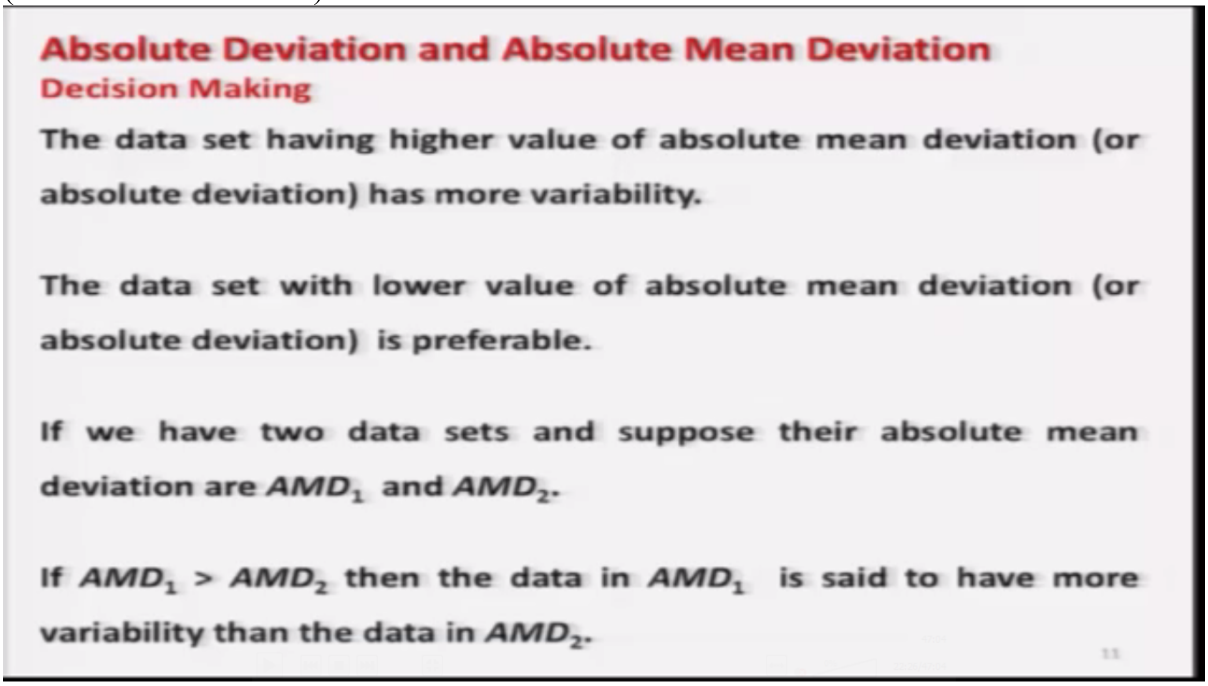
So this is called the absolute mean deviation.

(Refer Slide Time 21:48)



Now what are these absolute deviation or absolute mean deviation doing? So you can see here either it is absolute deviation or say absolute mean deviation. They are trying to present the information on the variability of the data in a comprehensive way. So this absolute mean deviation measure the spread and scatteredness of the data around, preferably the median value, in terms of absolute deviations.

(Refer Slide Time 22:26)

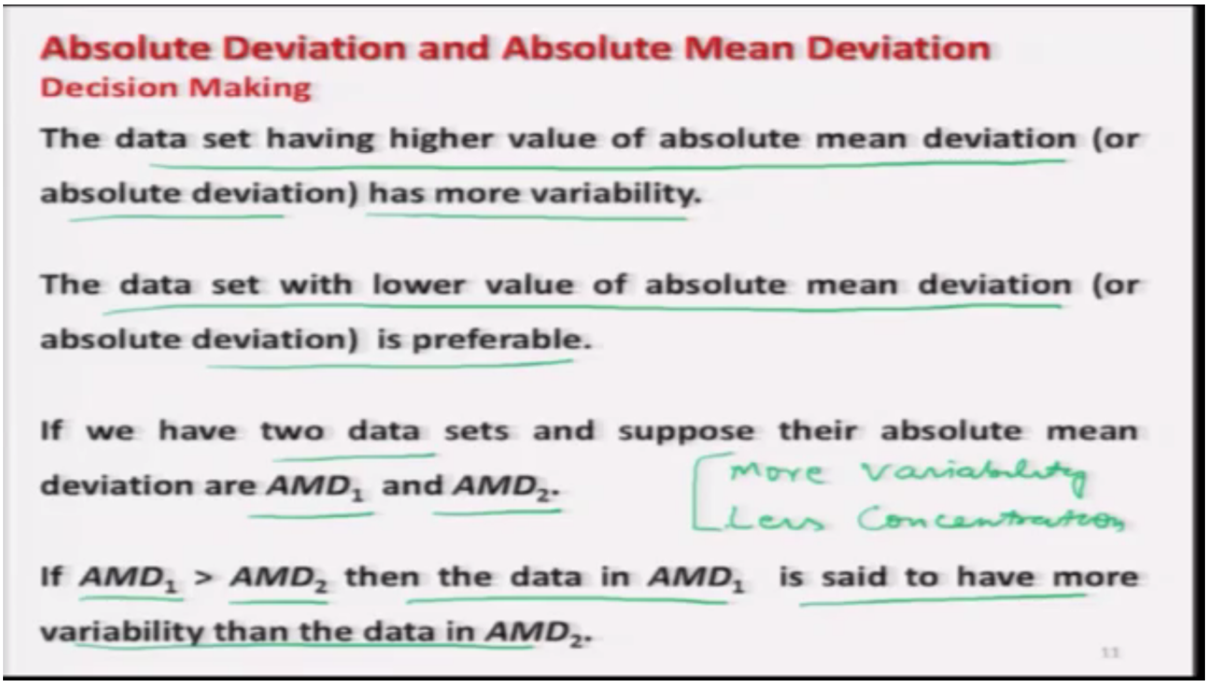


And in case if you want to know that how to make a decision season that if you have got more than one data sets and if you want to know that which dataset has got the more variability, then how to do it? So, again, we have a same rule as we had discussed in the case of interquartile range and quartile deviation.

So I will say here the data set which is having the higher value of absolute mean deviation or even the absolute deviation is said to have more variability and the data set with lower value of absolute mean deviation or the absolute deviation is preferable. And suppose we have two data sets and we try to compute the absolute mean deviation and we computed say AMD_1 and AMD_2. Then we say that if the value of AMD_1 is greater than the value of AMD_2, then the data in the AMD_1 is said to have more variability than the data in the AMD_2.

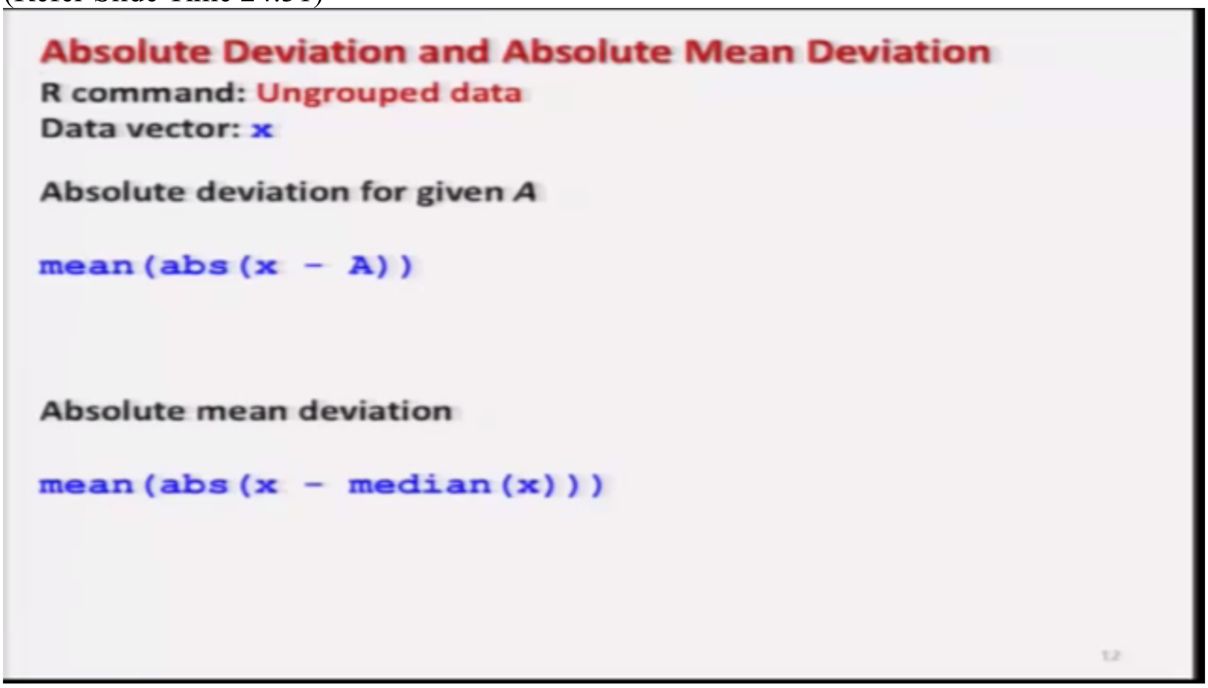
One thing what you have to keep in mind, there are two interpretation. More variability or second is less concentration. They have got the similar meaning.

(Refer Slide Time 23:51)



What I'm trying to say? More variability means that the data is more scattered around the value. When I say less concentration, that means the data is less concentrated around the median value. So, usually, people provide their inferences using that term variability or the concentration. So you have to be careful. More concentration means less variability. Okay.

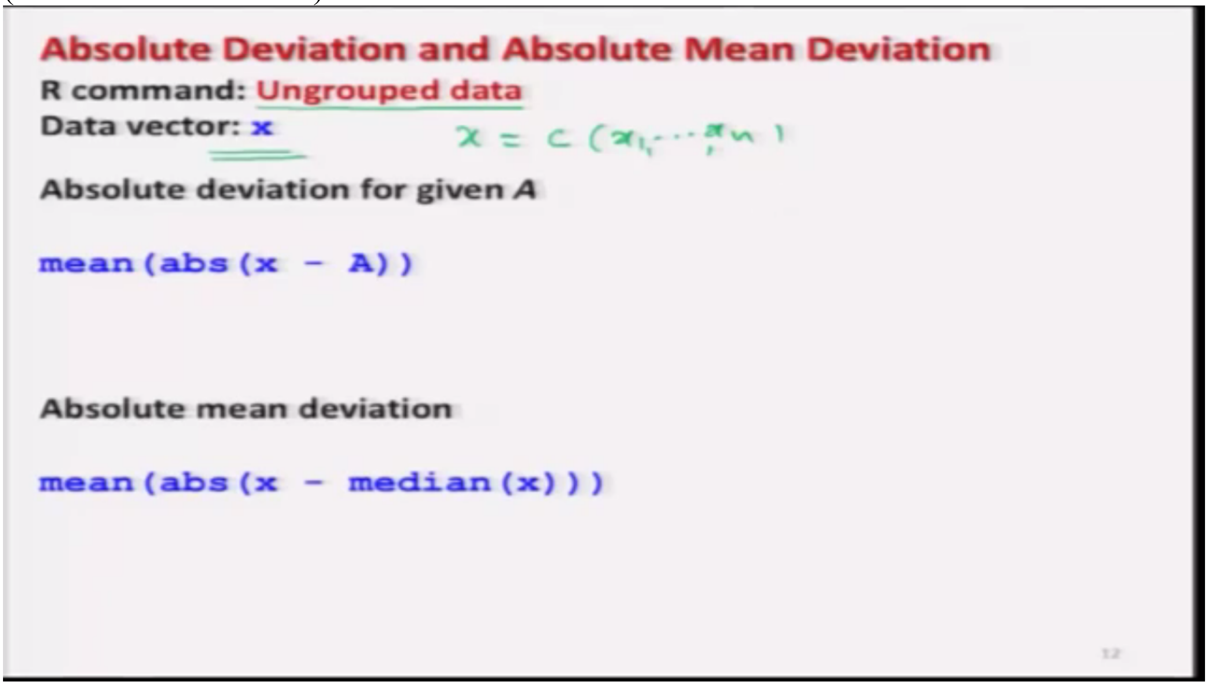
(Refer Slide Time 24:31)



Now the question is how to compute it on the basis of R Software? So now I have two cases for ungrouped data and second for grouped data. So in case of ungrouped data, I'm denoting the data vector by here x. So this is going to be x = c (x_1, x_2,...,x_n) separated by comma.

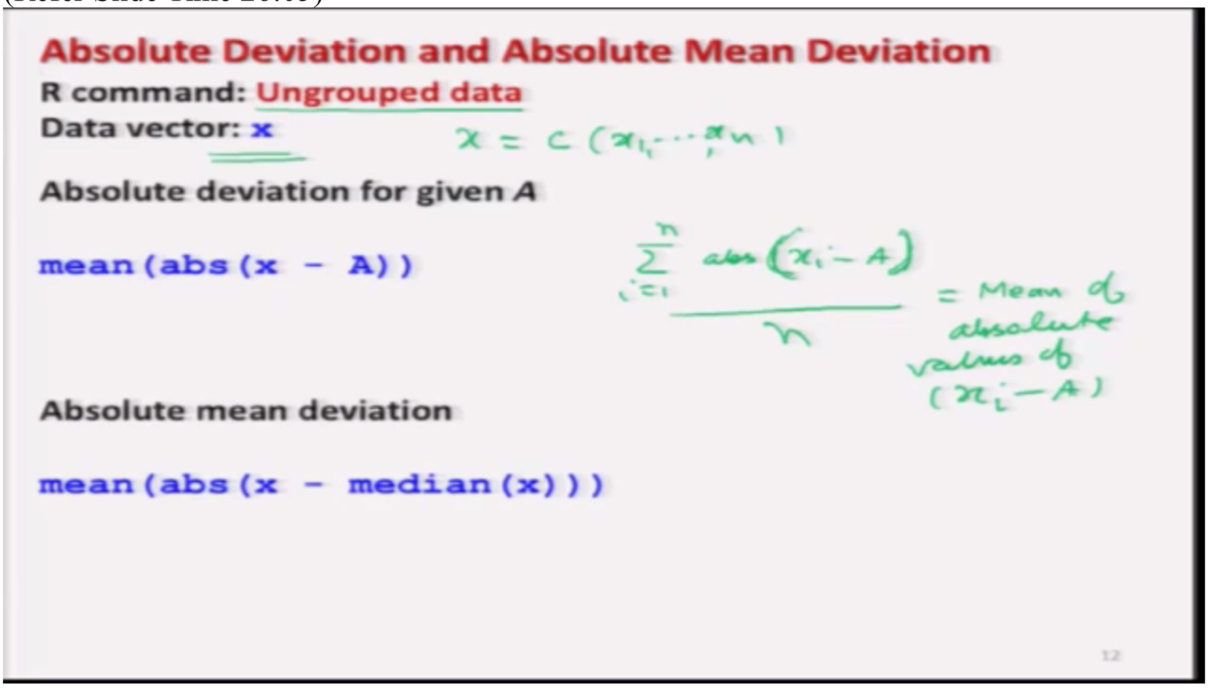
Now, once again, I will try to address that in R there is no built-in function to compute the absolute deviation or the absolute mean deviation, but by using the built-in commands inside the base package of R, it is not difficult at all to write down the command for absolute deviation or for the absolute mean deviation. Let us see how.

(Refer Slide Time 25:24)



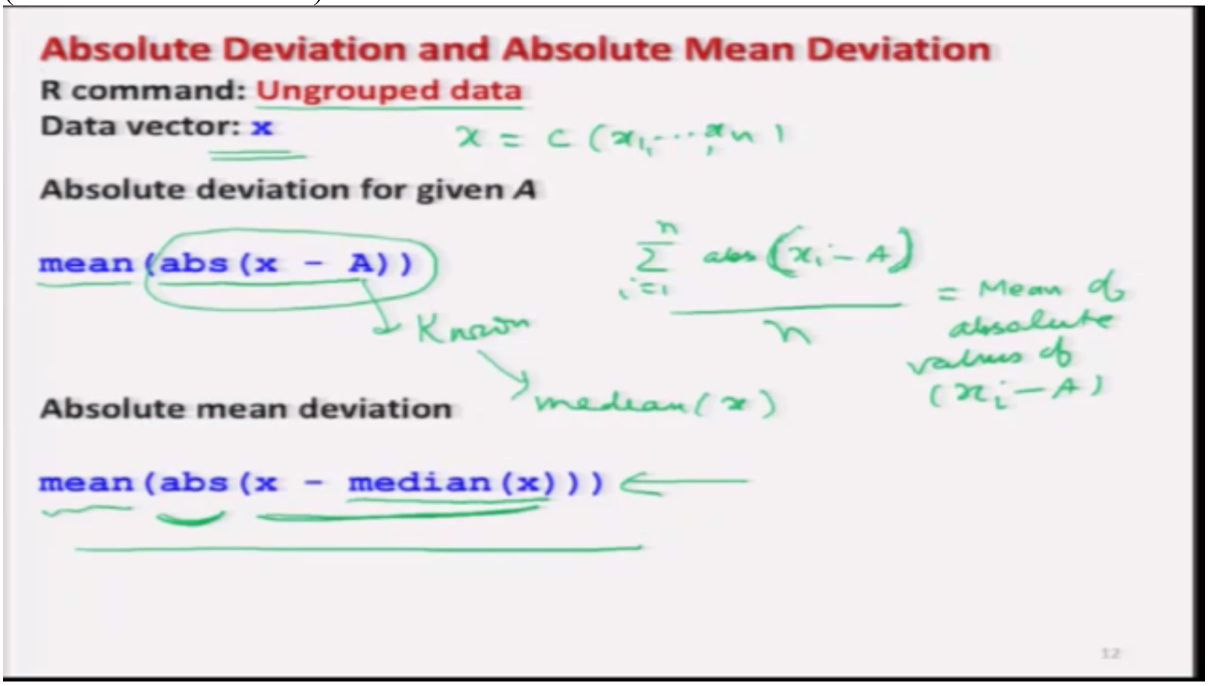
If you try to see what I'm trying to do, I have the observations absolute value x_i - A and I'm trying to find out the absolute value of this x_i - A. So for that I have the function here absolute abs (x_i - A), and then I'm trying to find out the sum of these and divided by the number of observation, which is nothing but the mean of absolute values of x_i - A.

(Refer Slide Time 26:05)



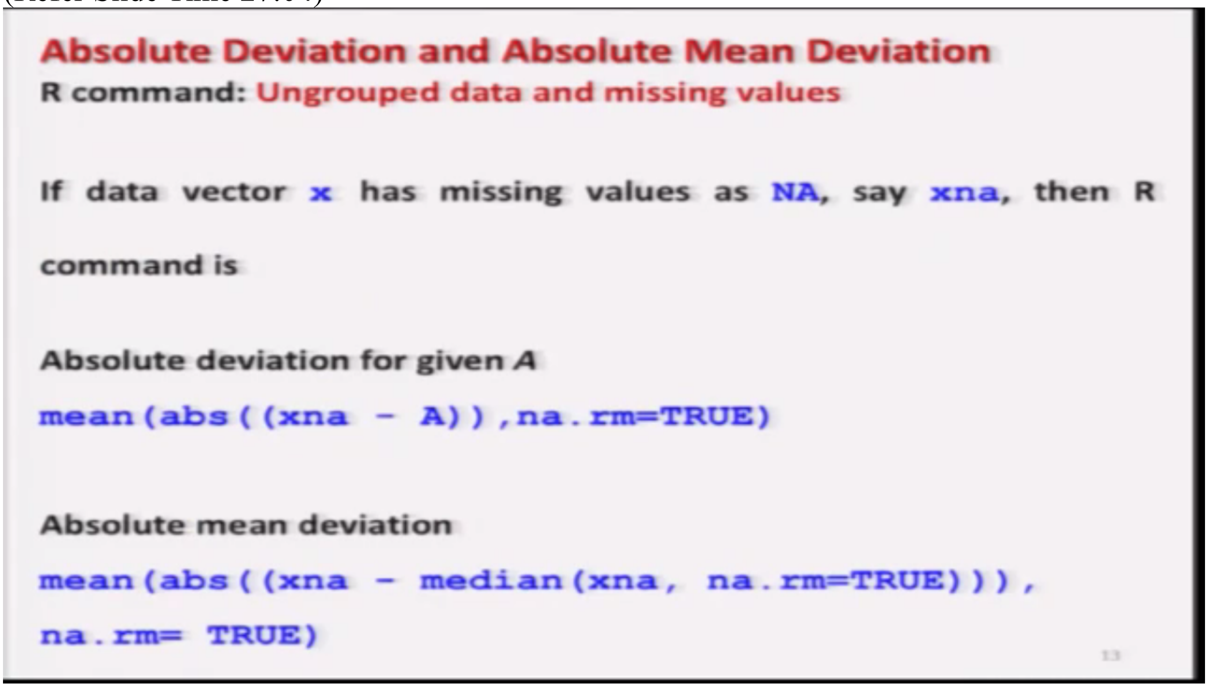
So I have used this concept here, and so, first, I try to write down the absolute value of all the data vectors around a given point A and then I try to find out the mean of all such absolute values. So by writing mean inside the argument absolute (x - A) we will find out the value of the absolute deviation around any given value A. So A is assumed to be here known and if I try to replace A by median of here x, then I know how to compute the median. The command here is median of the data vector x. So I try to find out the absolute value of the deviations x and median(x) by using the command abs and then I try to find out their arithmetic mean.

(Refer Slide Time 26:58)



So this command will give me the value of absolute mean deviation.

(Refer Slide Time 27:04)



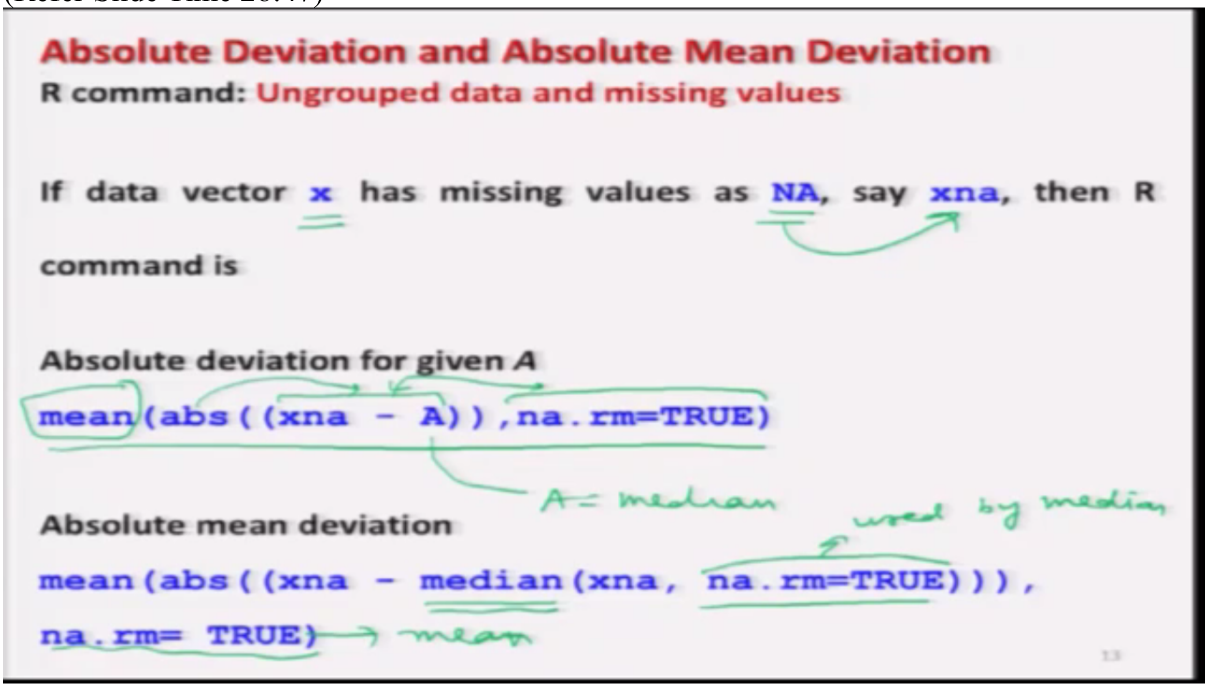
Now in case if I have ungrouped data or the missing data, so what I say is suppose I have this data vector here x and whatever are the values missing in this vector, they are denoted by NA and those values are collected inside the data vector xna. So in case if I have missing values in an ungrouped data, then absolute deviation is going to be computed by this command.

You see I have done nothing here. If you have understood the command that how to write or how to handle the missing value, I have written the same command earlier, but I have added here the command na.rm = TRUE.

So now this is telling that please try to compute the values of xna - A after ignoring the missing values using the command abs and then whatever is the outcome, operate the command mean on it. And similarly, in case if you want to find out the absolute mean deviation in such a case, simply try to replace A by median(x).

So in the same command, I'm trying to simply replace A by the median, but remember one thing that median also has to be computed using the command na.rm = TRUE. And then, first, na.rm, this is going to be used by the command median and this na.rm, this is going to be used by the command mean. So you simply have to be little bit careful while writing this expression. Sometime we make a mistake of the this brackets. Okay.

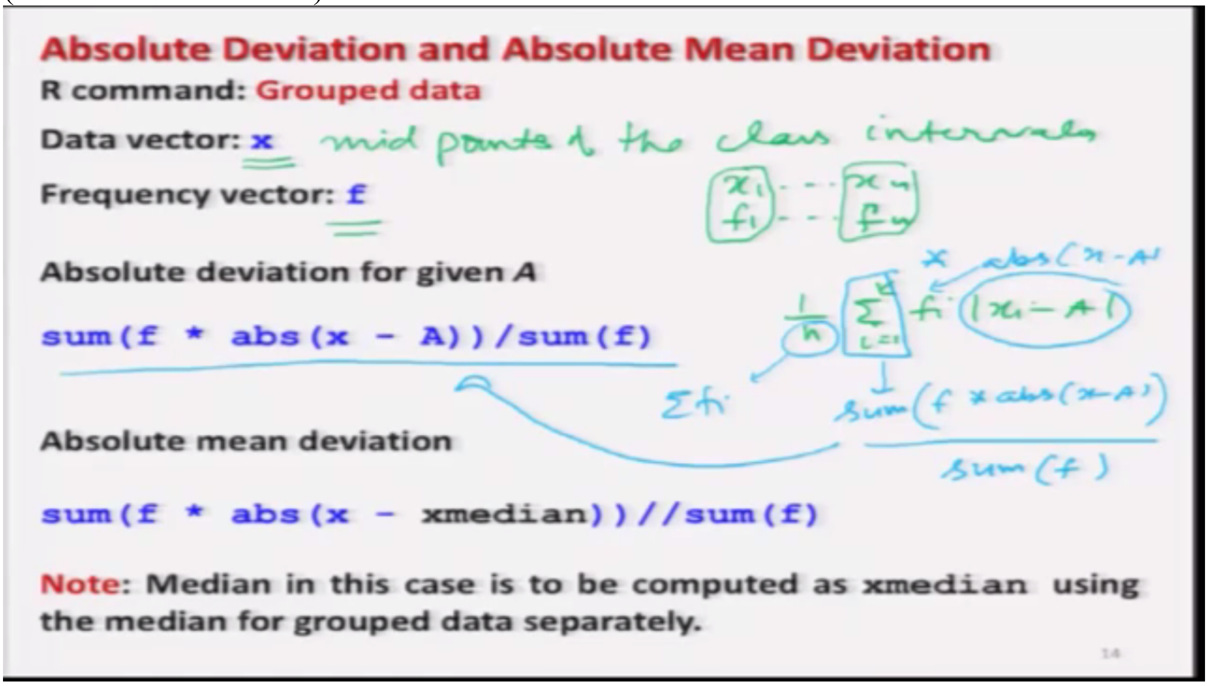
(Refer Slide Time 28:47)



Now, similarly, when we have grouped data, then now I will assume that my data vector here is x, which is essentially denoting the midpoints of the class intervals, and this data has the frequency f. So $x_1, x_2,...,x_n$, they are my class intervals with frequency f_1, f_2, see here f_n. Right.

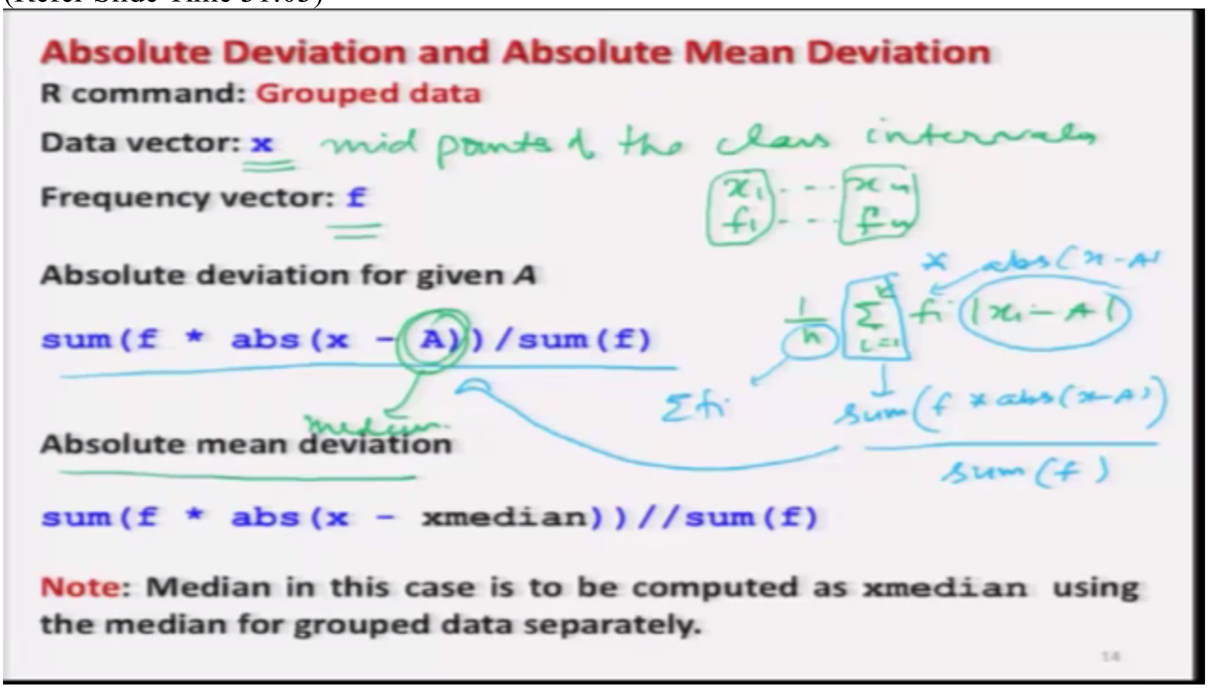
Now in this case, the absolute deviation is simply going to be the 1 upon n summation i goes from 1 to K and absolute f_i into absolute value of x_i - A. So I try to compute this x_i - A by the function absolute(x_i - A) and then I have to multiply by the corresponding f_i. So I can use here the operator * and then I have to find out here the sum. So I can say here first of all find down the absolute value of x - A. Then multiply it by here vector f and then whatever is the command here, this you have to find the sum, and it has to be divided by here n where this n is going to be the Σf_i. So I can divide it by here sum by say here f. Right. And the same thing has been written here in this command here.

(Refer Slide Time 30:40)



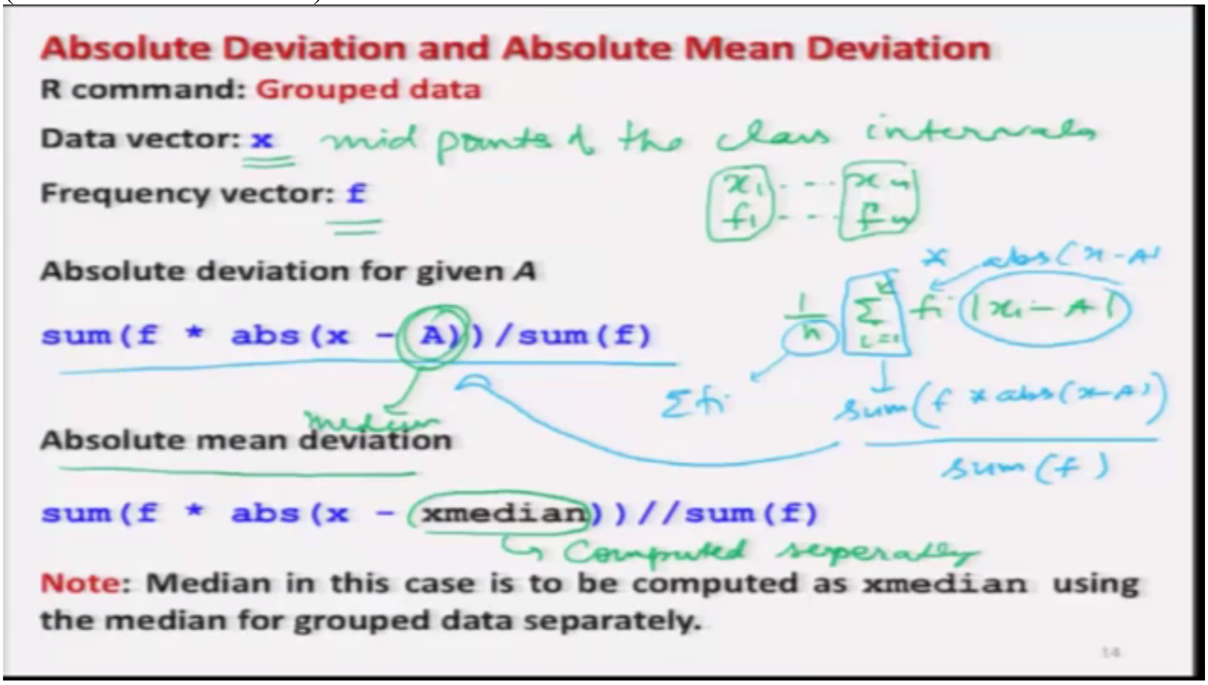
So once again you can see here that there is no built-in function, but you have to be careful, and you can easily write such commands. And similarly, if you want to compute the absolute mean deviation, what I have to do? I simply have to replace this A by here median.

(Refer Slide Time 31:03)



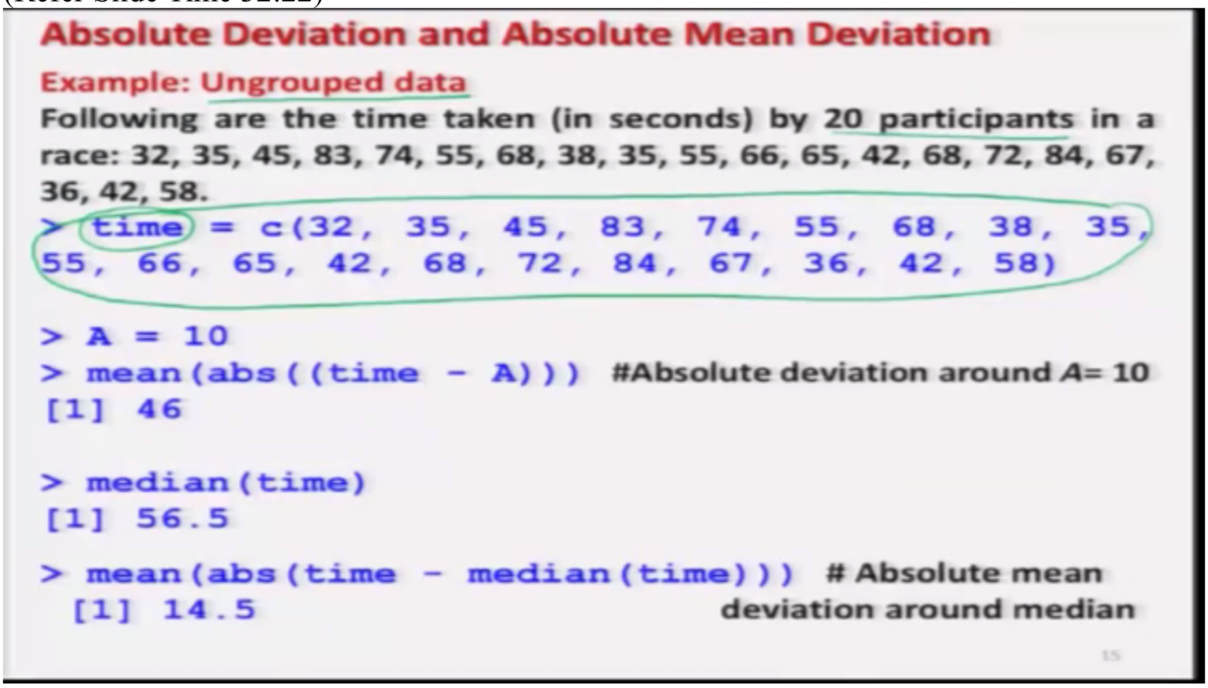
So just in order to make it clear that here we are not going to use the command median directly, but we need to compute the value of the median separately, I am using here a black colour font and I'm denoting here as xmedian so that you remember that this value has to be computed separately using the functions and command that we have used in the case of computation of median in case of grouped data. Right. Okay.

(Refer Slide Time 31:48)



So now let me take here an example and I will try to take the same example. So I will try to first treat the data set as ungrouped data set and then I would convert it into an ungrouped data set. So, first, I try to consider the case of ungrouped data set. You have the same example that we have used earlier that we have the data of 20 participants, and this data has been recorded inside a variable say a time, and this is the data vector here.

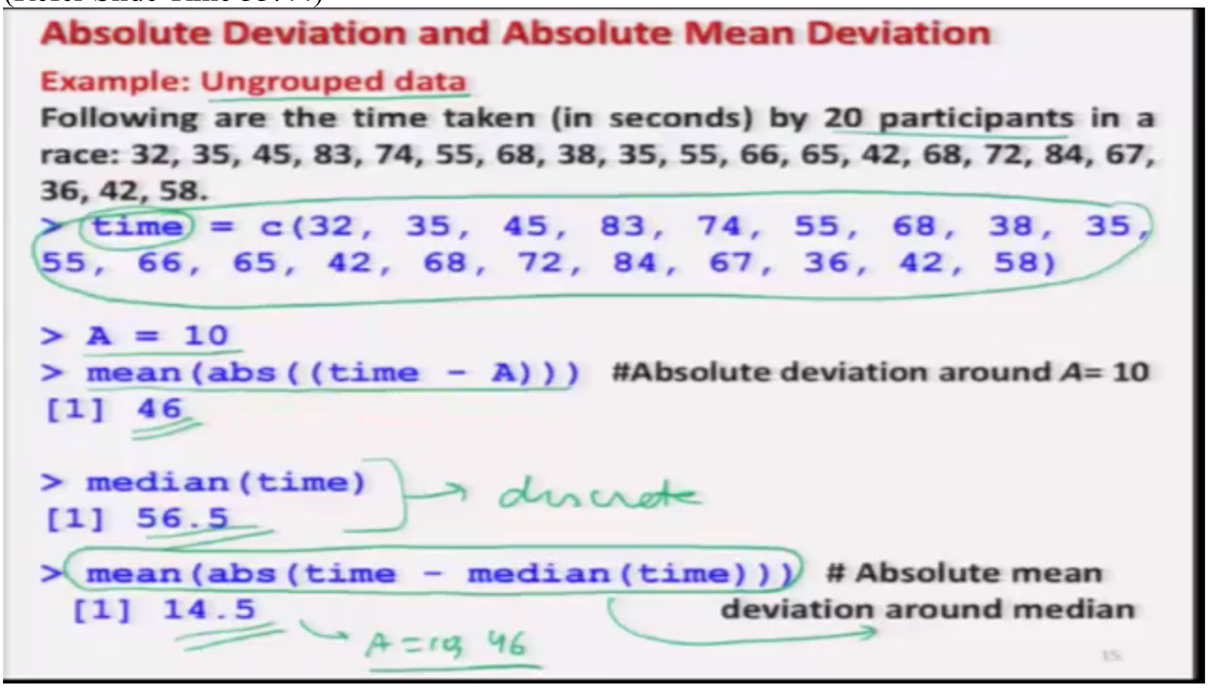
(Refer Slide Time 32:22)



And in order to compute the absolute deviation, I try to choose say value of A. Just for the sake of understanding, I'm trying to choose here the value of A to be 10. And now once I operate this command what we have just noted mean of absolute value of time - A, this will give me the value here 46. And if I try to find out the median of the time, you can see here, this value will come out to be 46.5 because this is the discrete variable. You have the ungrouped data. So you can use this command, and now using this command directly inside the command here, I'm trying to compute the absolute value of these deviations of time from the median and then I'm trying to find out the mean. So this is giving me the value of absolute mean deviation around median. So this value comes out to be here 14.5.

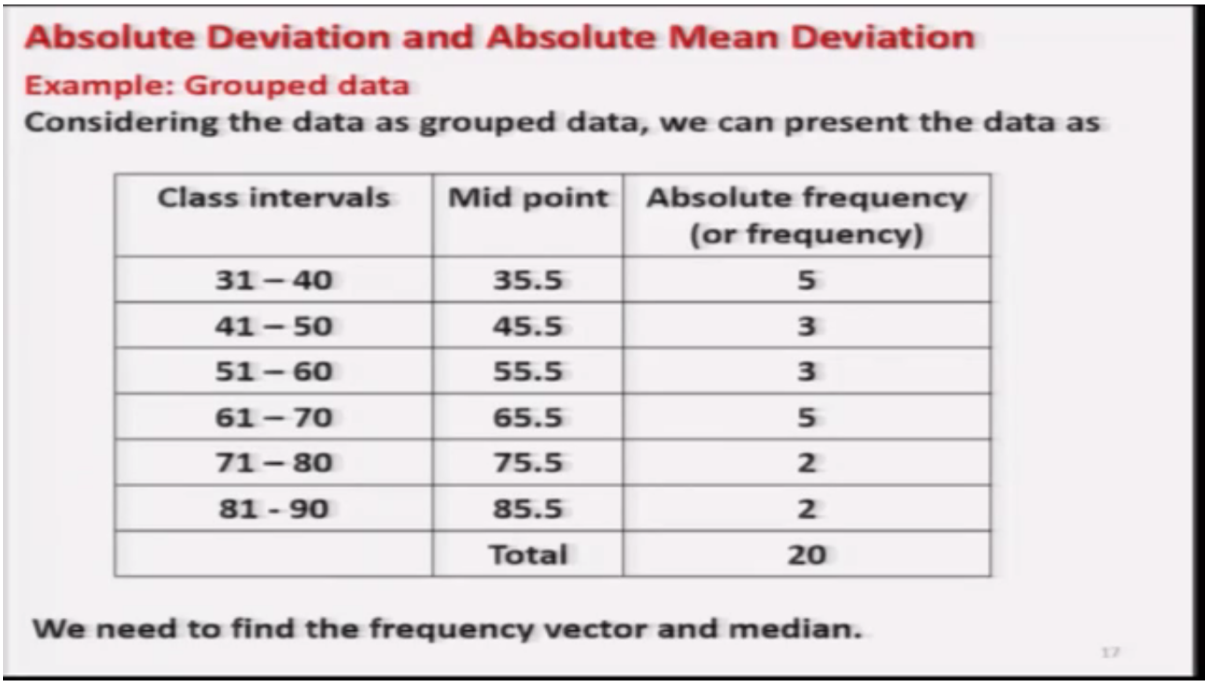
So as we had discussed that this absolute mean deviation is going to be minimum when it is measured around the median, so you can see here when I try to choose here this value to be here equal to 10, then this value was coming out to be 46, but now this value is coming out to be 14.5.

(Refer Slide Time 33:44)



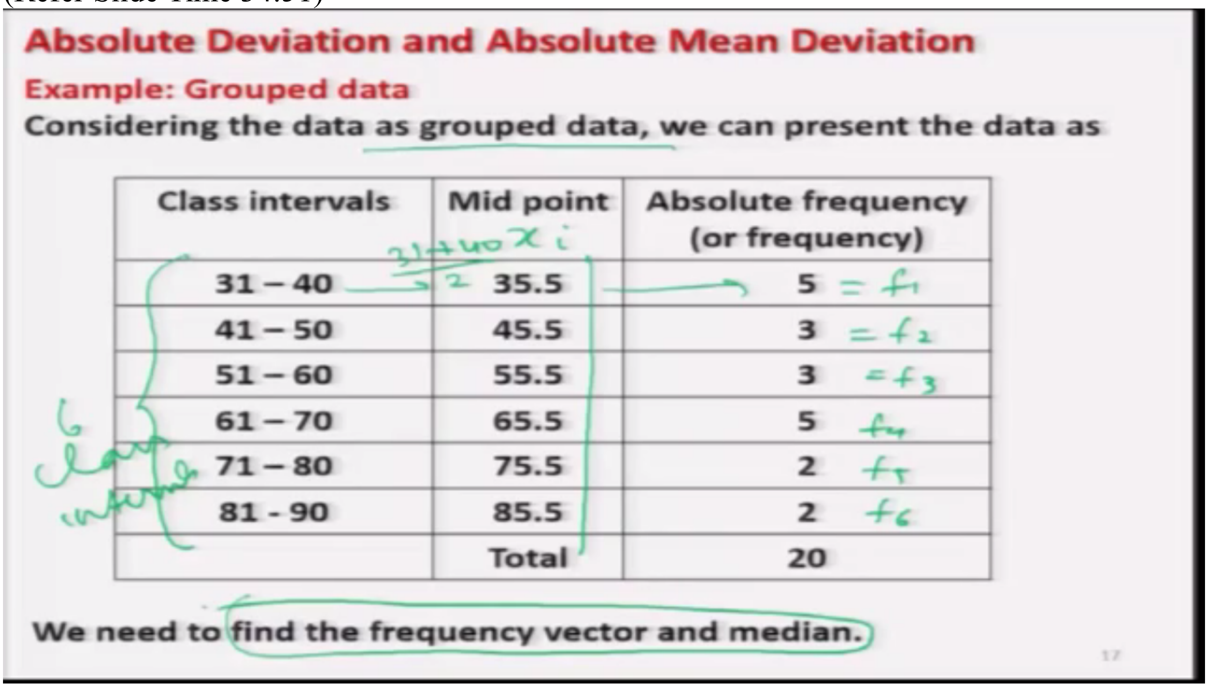
Okay. Before I try to go to the R console, let me show you here the screenshot of the same operation that we are going to do, but before that, let me consider this data as a grouped data and show you how to compute the median in this case also.

(Refer Slide Time 34:00)



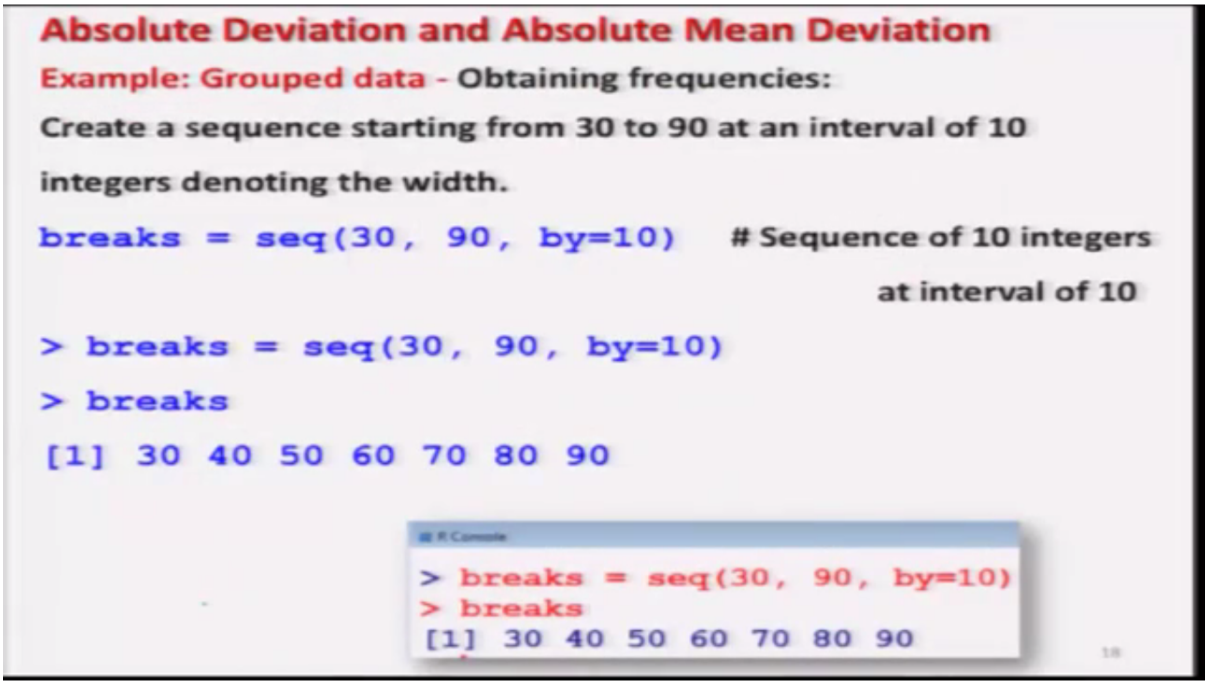
So now considering the same dataset as a grouped data, I have divided the entire data into six class interval and I have found their mid-value x_i here in the second column. So you can see here this value is 31+40 divided by 2. So this is 71 by 2 which is 35.5 and so on other values and here are the frequencies. So 5 is the frequency of class 1; f_2, which is the frequency of second class, is 3; then f_3; then f_4; then f_5 and then f_6.

(Refer Slide Time 34:51)



Now what is our objective? We simply want to find out the frequency vector and median from this given set of data.

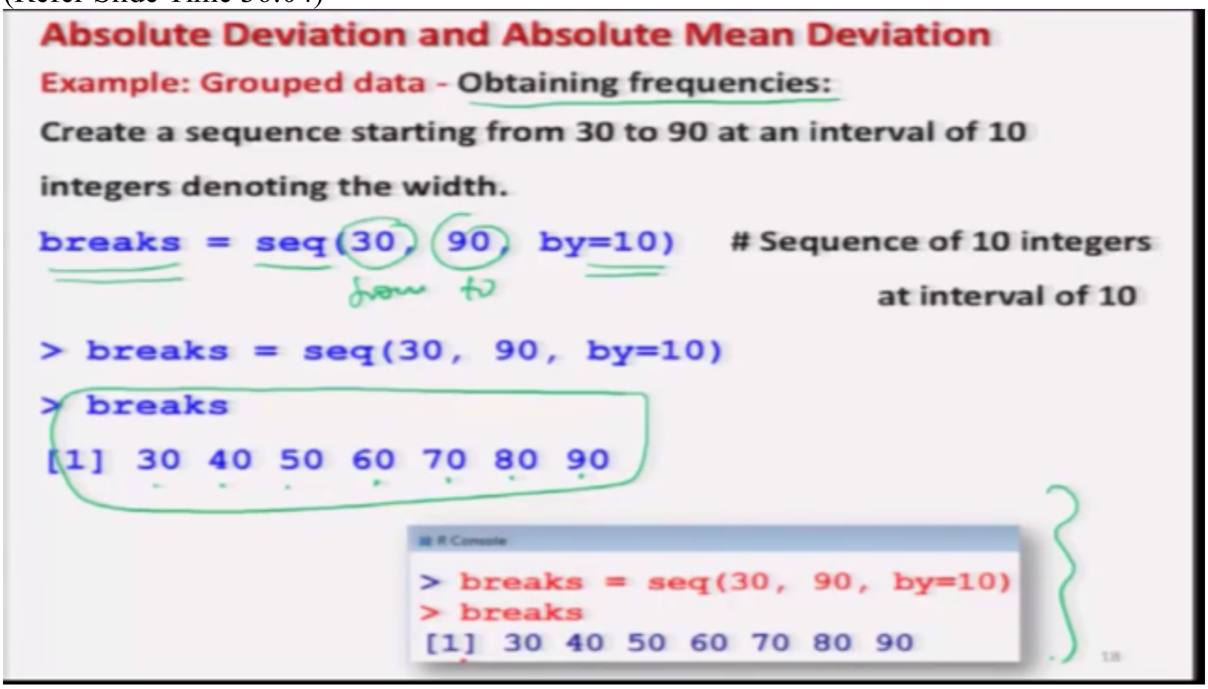
(Refer Slide Time 34:53)



Now how to find out the frequency vector and how to find out the median? I'm not going to explain here in detail because I already have discussed it in more detail when I was trying to compute the median and I also have used it in the earlier lectures. So what here just for your understanding and so that you can recall, I will simply try to give you the broad steps and these broad steps are simply I have taken from the slides that I had used earlier.

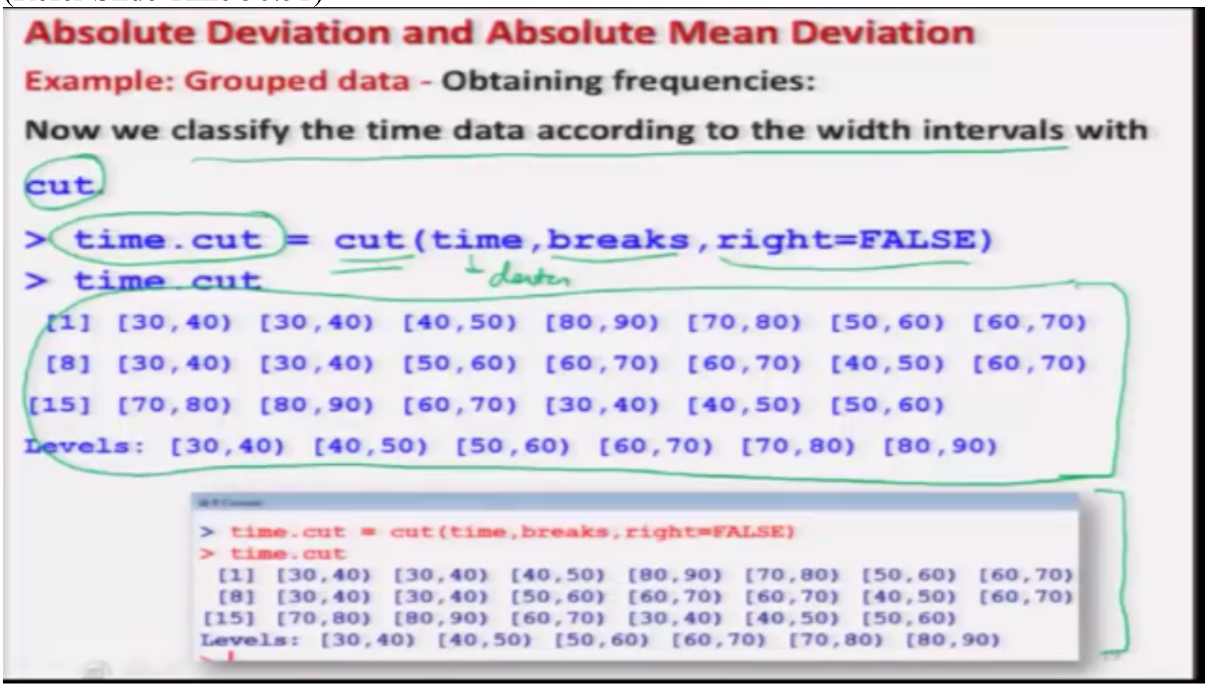
So my first objective is this. I want to obtain the frequency vector. So we had used the command sequence seq to generate a sequence from 30 to 90 at an interval of 10 and we had denoted this by breaks, and this breaks has an outcome like 30, 40, 50, 60, 70, 80, 90 and this is here the screenshot. We have done it earlier.

(Refer Slide Time 36:04)



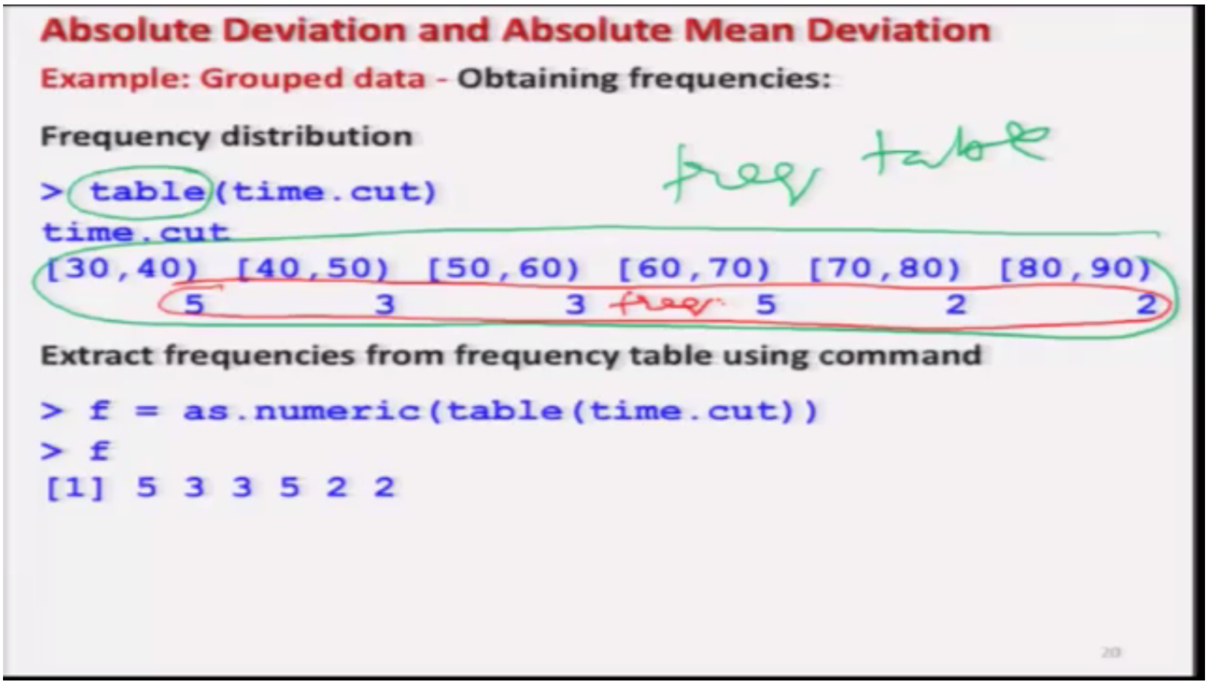
And then after this we wanted to obtain the frequencies. So in order to obtain the frequencies, first, we need to classify the time data in the class interval using the width that was defined by the breaks. So for that we have used the command cut and this was operated. The cut command was operated over the data vector on time using the breaks and we had used the option right = FALSE so that the intervals are open on the right-hand side and I had stored this, these values in time.cut vector, and this, these values were obtained here like this and this is the here the screenshot.

(Refer Slide Time 36:51)



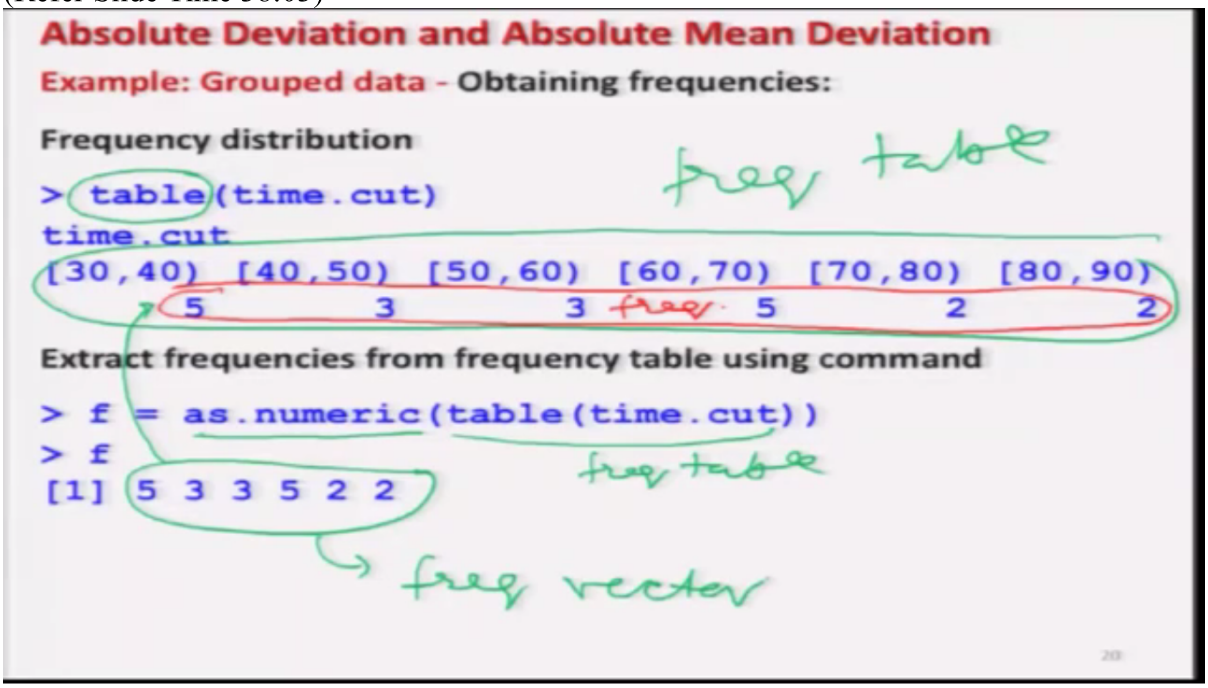
And here in this case if you recall the last row, which is indicated with levels, that is going to indicate the class intervals. Now after this we had used the command table to tabulate the values of time.cut and the frequency table was obtained here like this. This was the frequency table and the second row over here, 5, 3, 3, 5, 2, 2, this is going to give us the frequency values.

(Refer Slide Time 37:32)



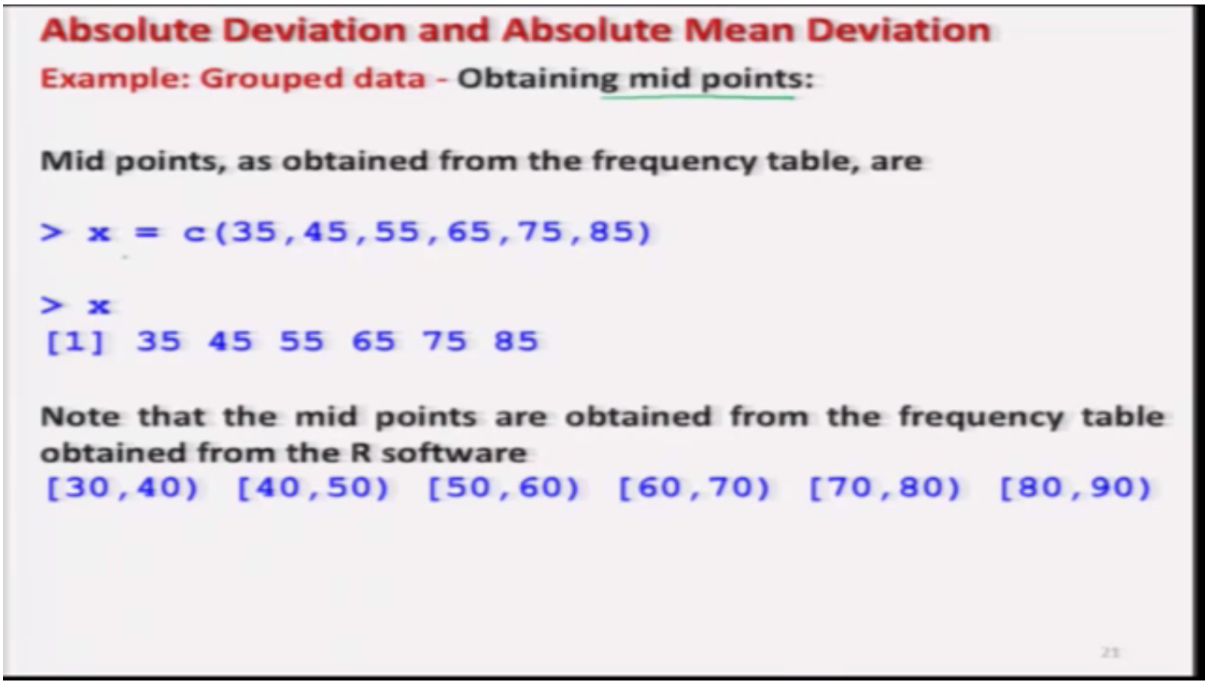
So we had obtained the frequency vector from this table using the command as.numeric and inside the arguments the name of the frequency table, and this gives me the here an outcome 5 3 3 5 2 2 and this outcome is the same as given here in this vector. So this is our frequency vector. So now this is our frequency vector.

(Refer Slide Time 38:03)



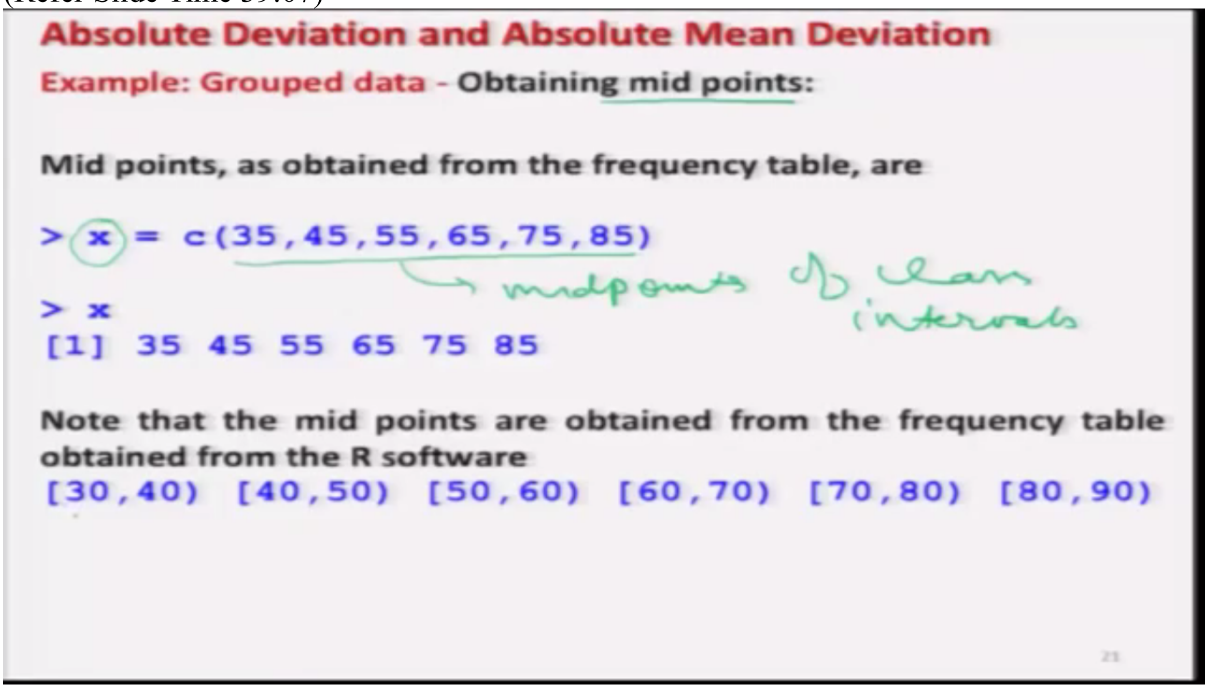
Now we need to find out the vectors of midpoints. So now how to obtain it?

(Refer Slide Time 38:12)



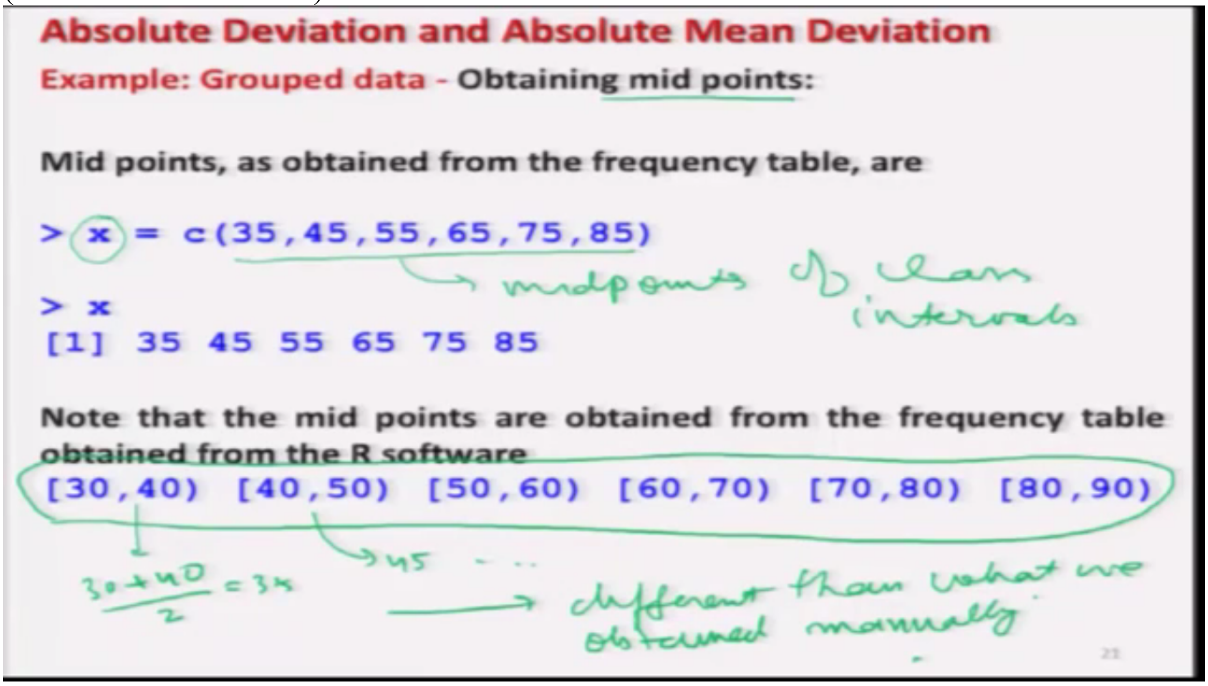
We had defined this vector x to be here the midpoints of the class interval. So these points are the midpoints of class intervals. How they are obtained? You see, one thing you have to be very careful here. When we are compiling the data and when we are trying to compile the data through the R Software using the command table, then in this case I have to use the midpoints of the class interval and the frequencies, which are provided by the R Software. Right.

(Refer Slide Time 39:07)



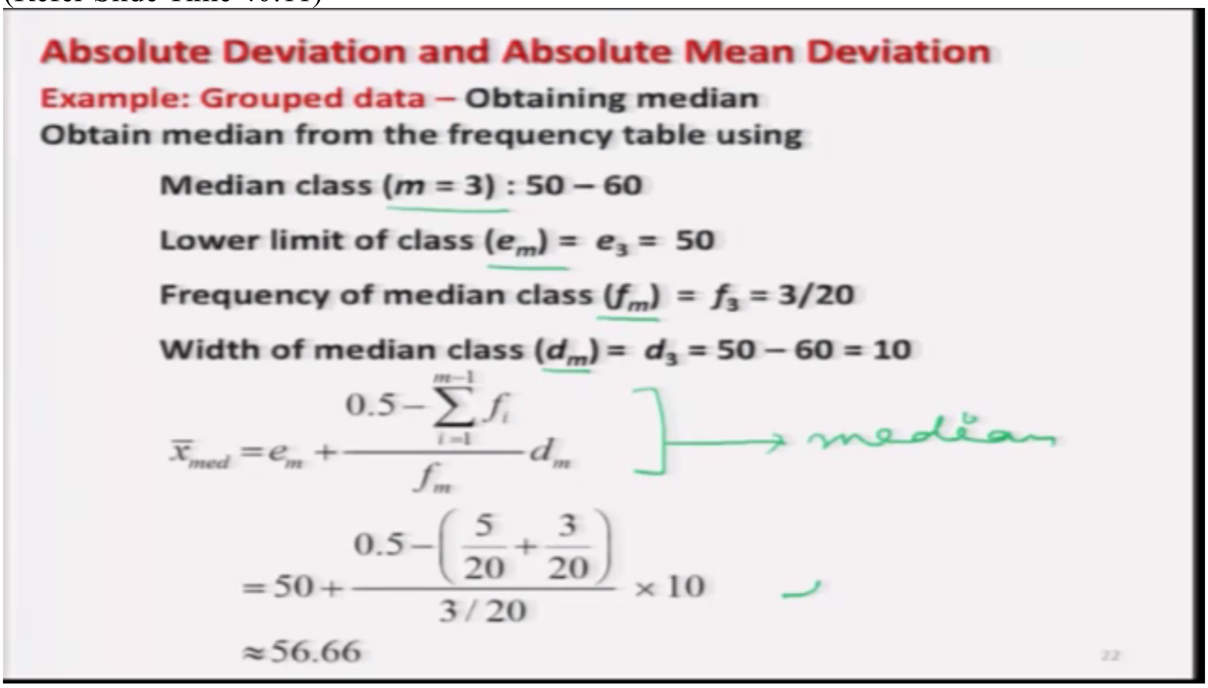
So here if you try to see in the earlier slide we had obtained the class intervals like this one. So the midpoint of this interval is (30+ 40) ÷ 2 which is equal to 35. The midpoint here is 45 and so on, and this is different than what we obtained earlier manually.

(Refer Slide Time 39:39)



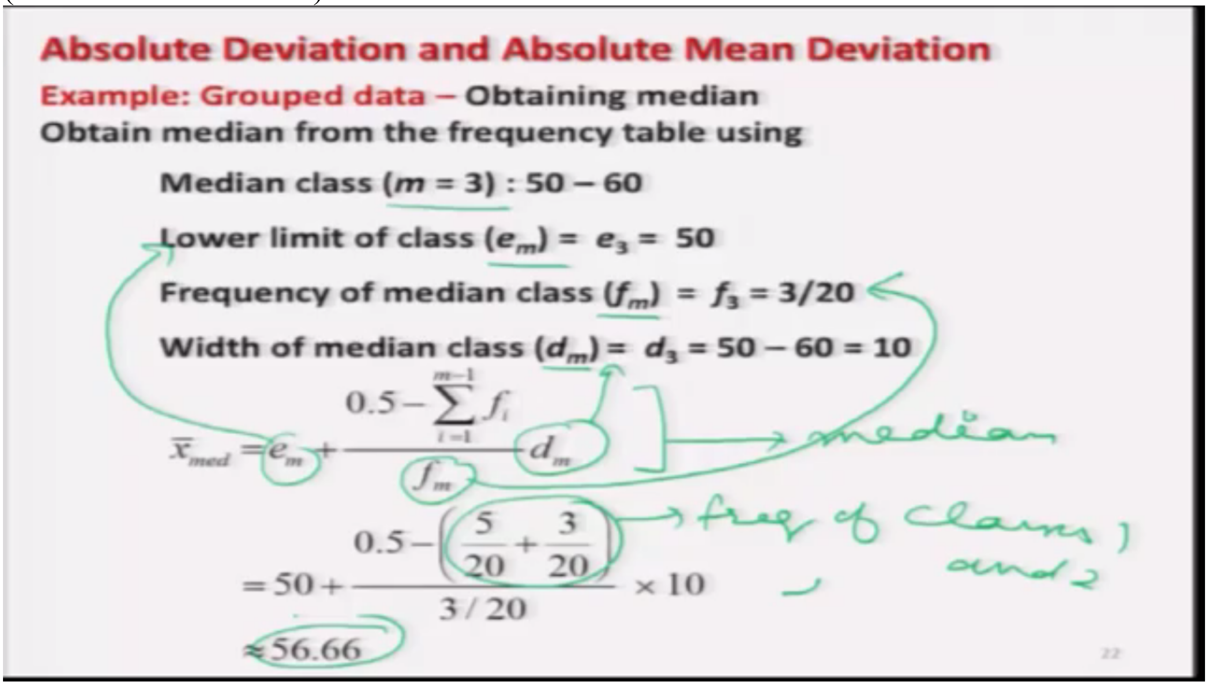
You may recall that this value was 35.5 and so on, but anyway we are now doing in the software and this will not make much different. So this is now here my vector here x. After this I have to use the expression for finding out the median. So if you remember we have used the notation m, e_m, f_m, d_m and we had this expression to find out the median, which we have discussed in the case when we had our lecture on median.

(Refer Slide Time 40:11)



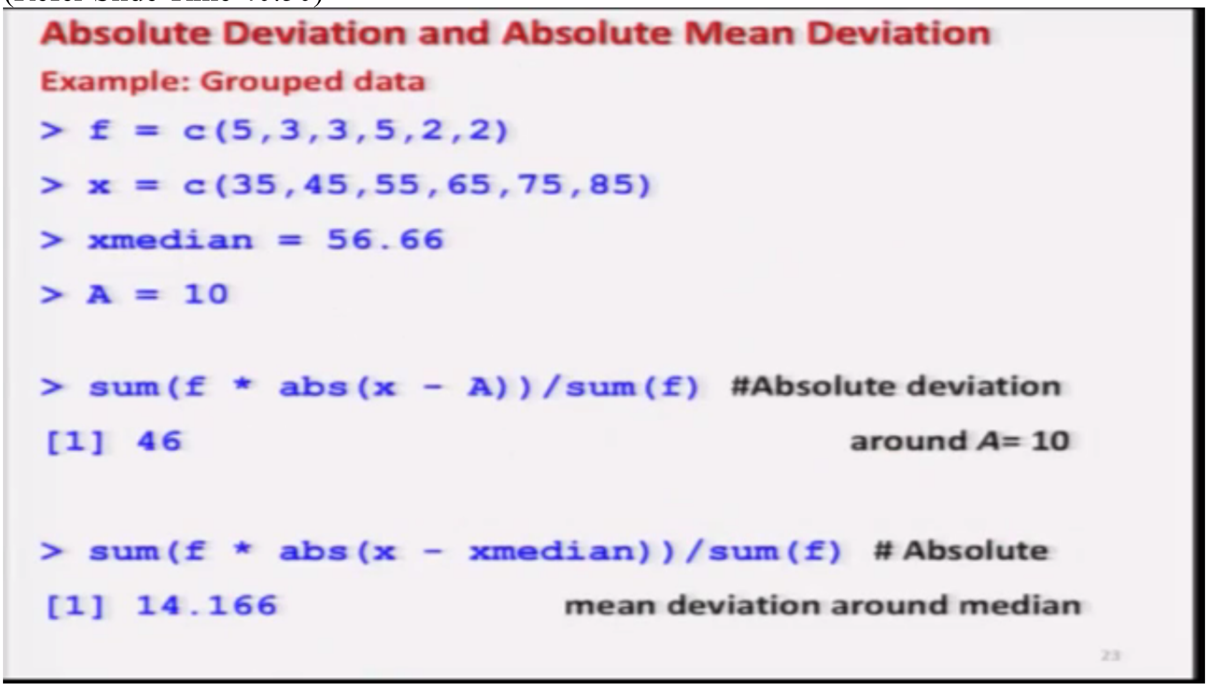
So now using this e_m, which is denoting the lower limit of the class, this f_m, f_m is going to give me the frequency of the median class, which is here, and some of these f_i's that is going to be the sum of first two classes frequency, frequency of classes one and two and d_m here is the width of the median class. This is 10. So now using these values over here, I get the value here 56.66.

(Refer Slide Time 40:45)



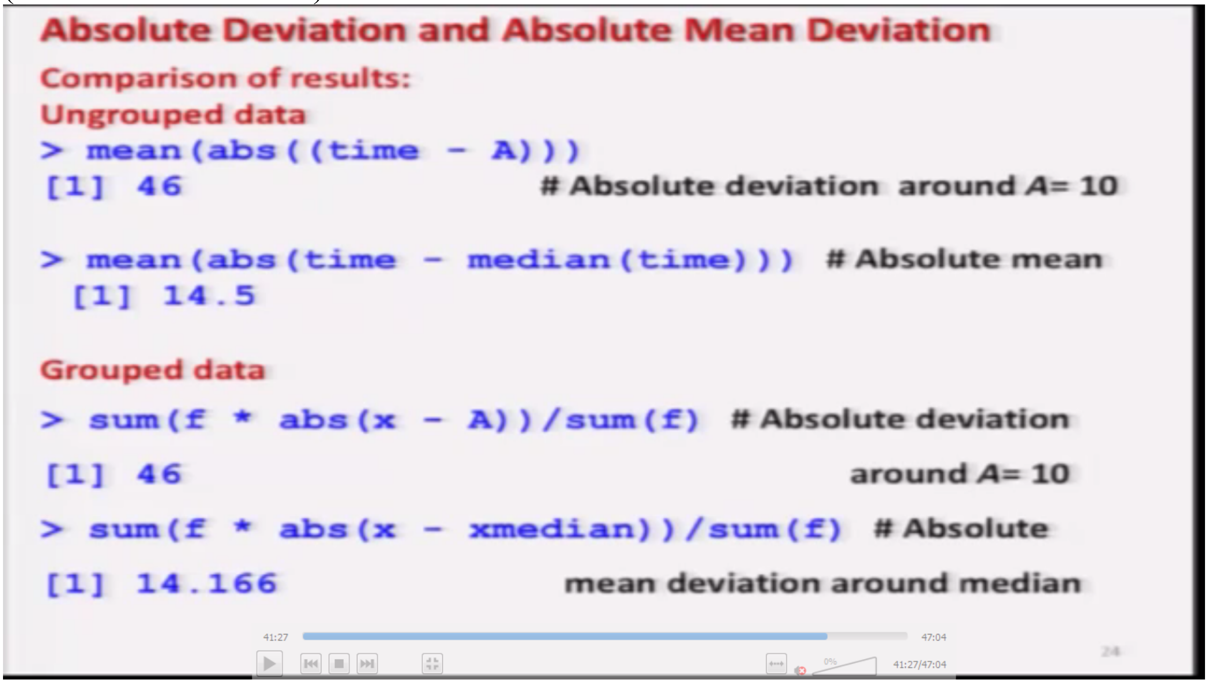
So what you have to notice here that in this case we have to find out the median separately.

(Refer Slide Time 40:50)



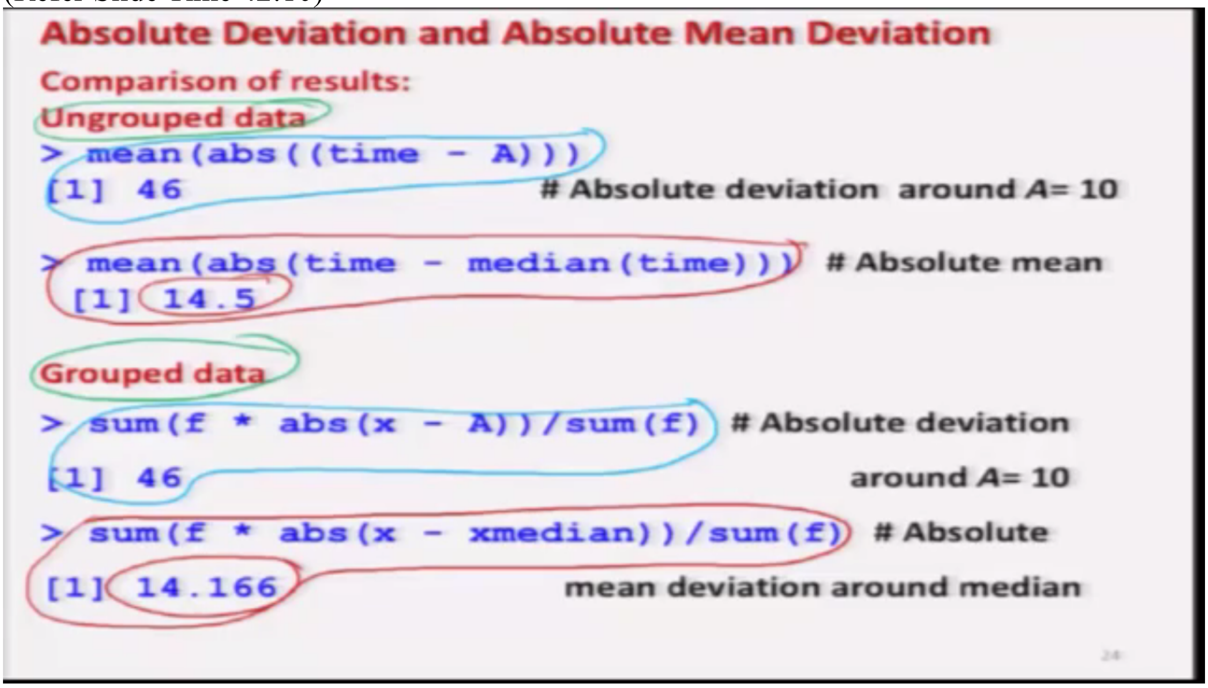
And now using these commands, now I have here the data vector on frequency f, data vector on see here x, which are the middle points, then the value of the median 56.6 and the value of here constant A, which I have chosen myself. So now using the same commands for finding out the absolute deviation with value A and finding out the absolute mean deviation around the median, but where the median has been defined separately, we get these values here.

(Refer Slide Time 41:26)



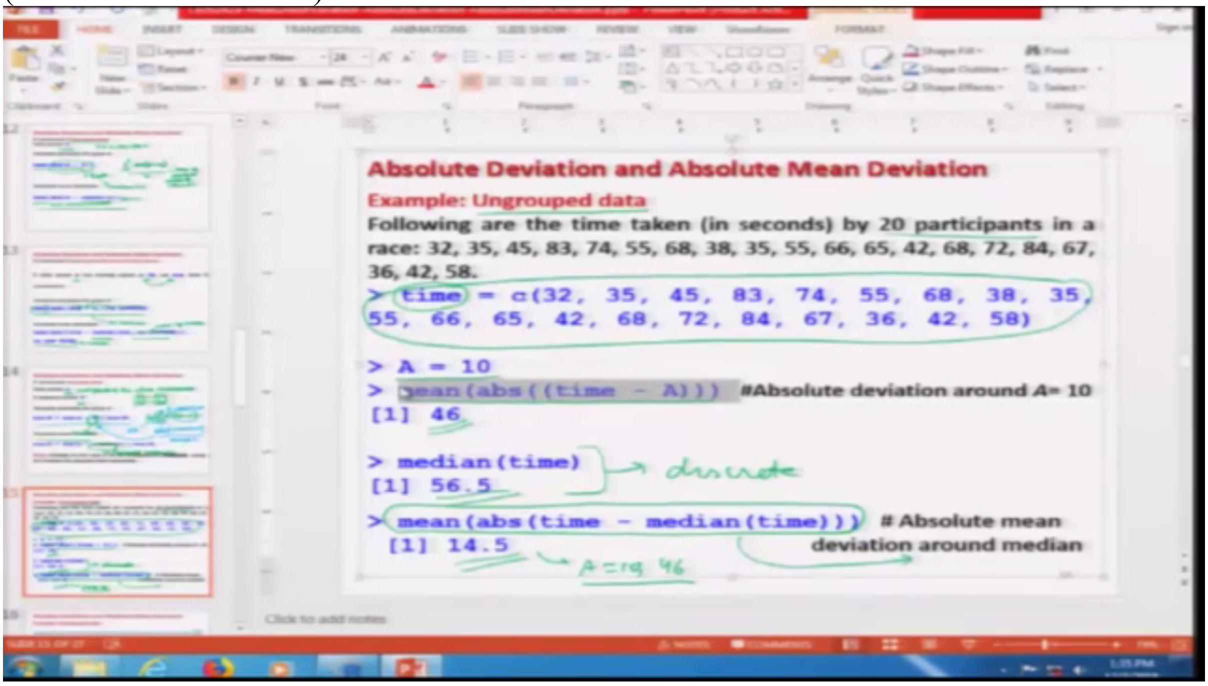
And in case if you try to compare the results that we have obtained for the grouped and ungrouped data, you can see there is not much difference actually, right? Here that this is for the ungrouped data and then for the grouped data. You can see here the values of this medians around A equal to 10, they are coming out to be the same and similarly the values which we have obtained for the absolute mean deviation in case of grouped and ungrouped data, they are not much different. This is 14.5 and 14.16, so 14.2, so there is not much a difference between the two here.

(Refer Slide Time 42:10)



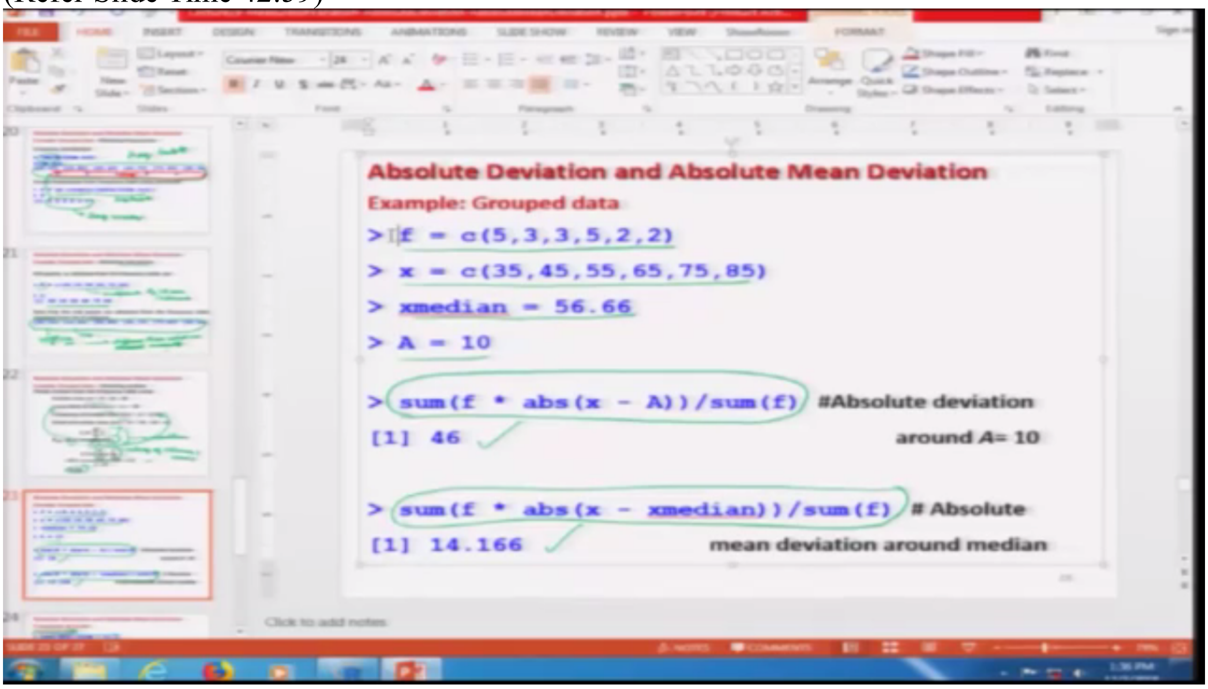
And now I would first try to show you these things on the R console. So we come to our example and where we try to find out the first the absolute deviation.

(Refer Slide Time 42:25)



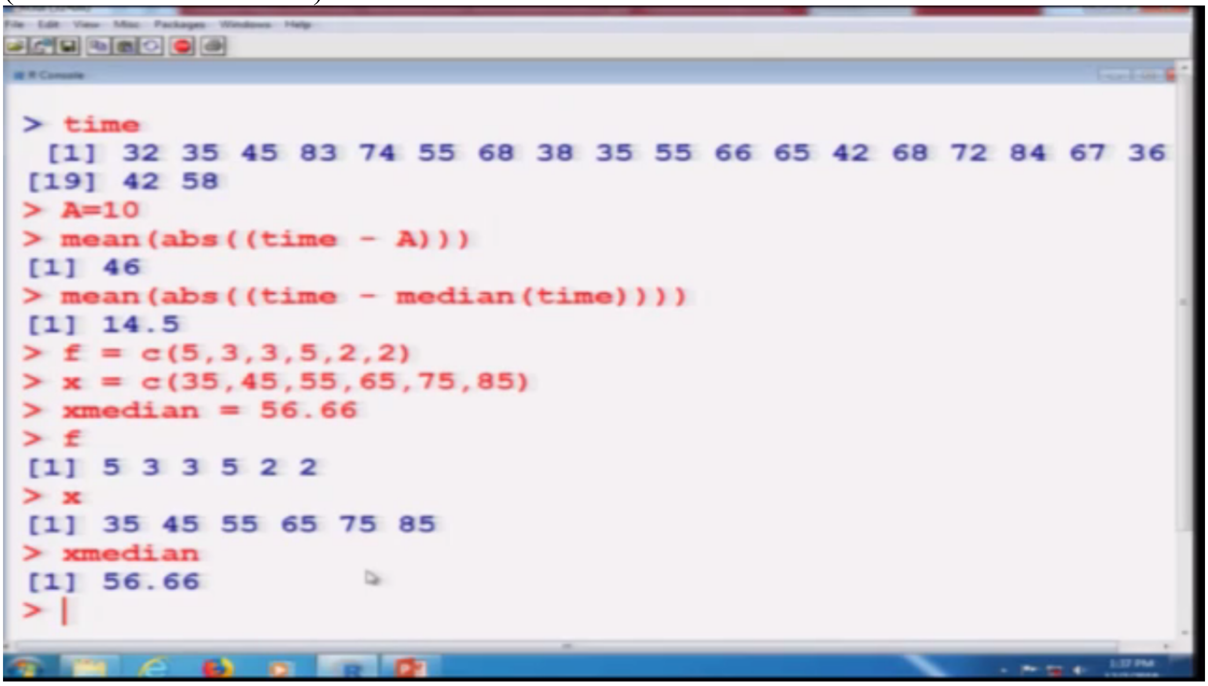
So I can show you that here I have the data here time, and I can take A to be here 10, and if I try to find out the absolute deviation, this is coming out to be 46, and if I try to replace A by median(time), this is coming to be here 14.5 and you can verify here that these are the same data set or the same result that we had obtained earlier.

(Refer Slide Time 42:59)



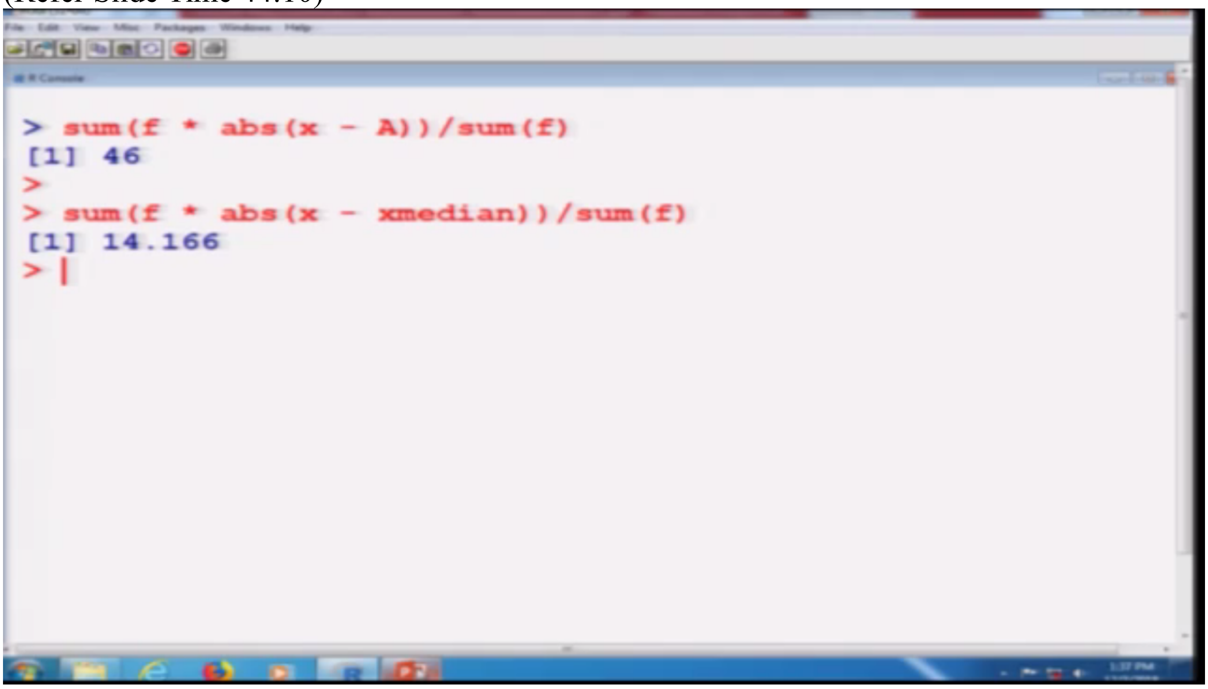
Now I try to come to the grouped data case where we have this, the frequency vector to be here like this. We have copied here and the vector of the midpoints which I am writing here like this and the value of here xmedian that has been obtained separately. This is here like this. So you can see here f is coming out to be here. X to be like here. xmedian is coming to be like as here. Right.

(Refer Slide Time 43:38)



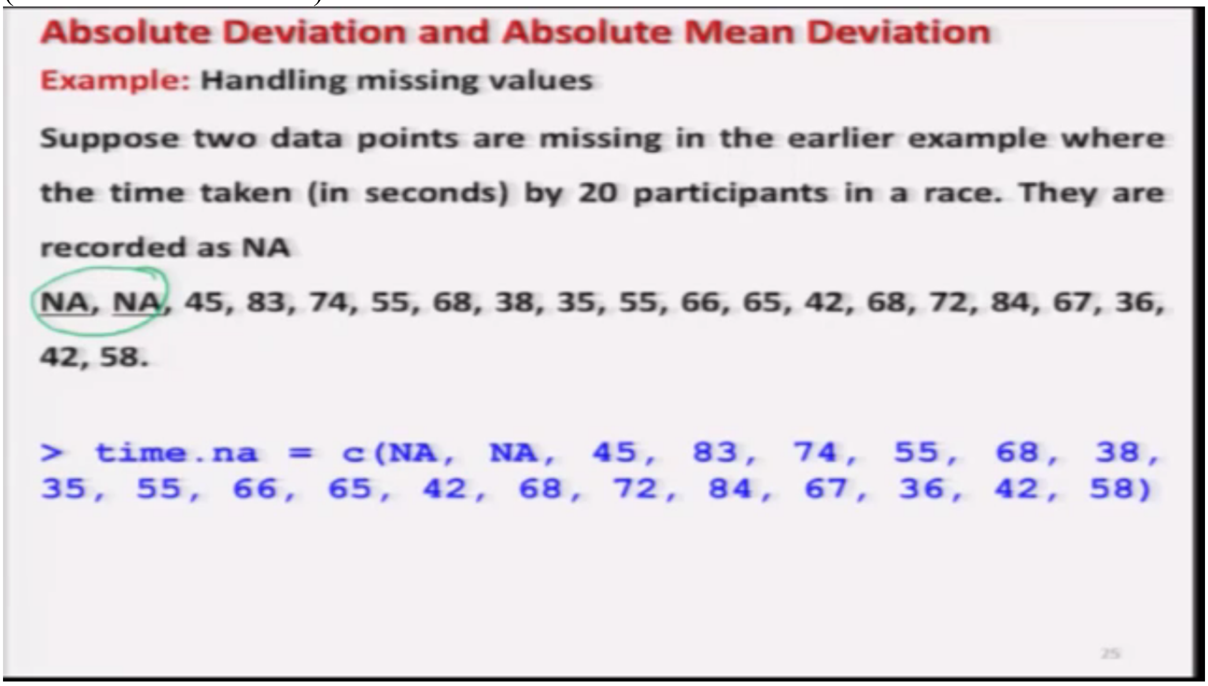
Now let me just clear the screen so that you can see clearly. Now using these values, I'm trying to compute the absolute deviation and absolute mean deviation. You can see here this is coming out to be 46 and similarly if I try to compute here the absolute mean deviation, this will come out to be like this. So you can see here that it is not very difficult to.

(Refer Slide Time 44:10)



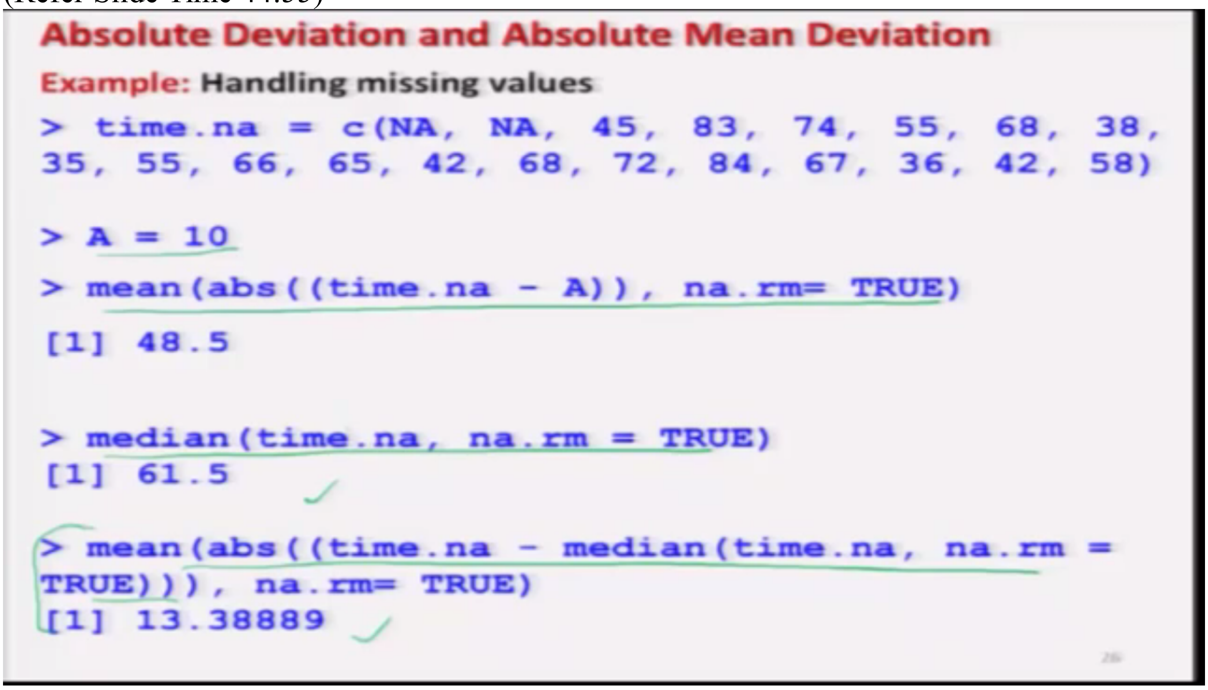
Now I will just quickly show you the same example that in case if you have got missing values in the data. So I will try to take the same data set, and I'm assuming that first two values are replaced by here NA and I will try to show you that how to compute these values using this missing value data.

(Refer Slide Time 44:27)



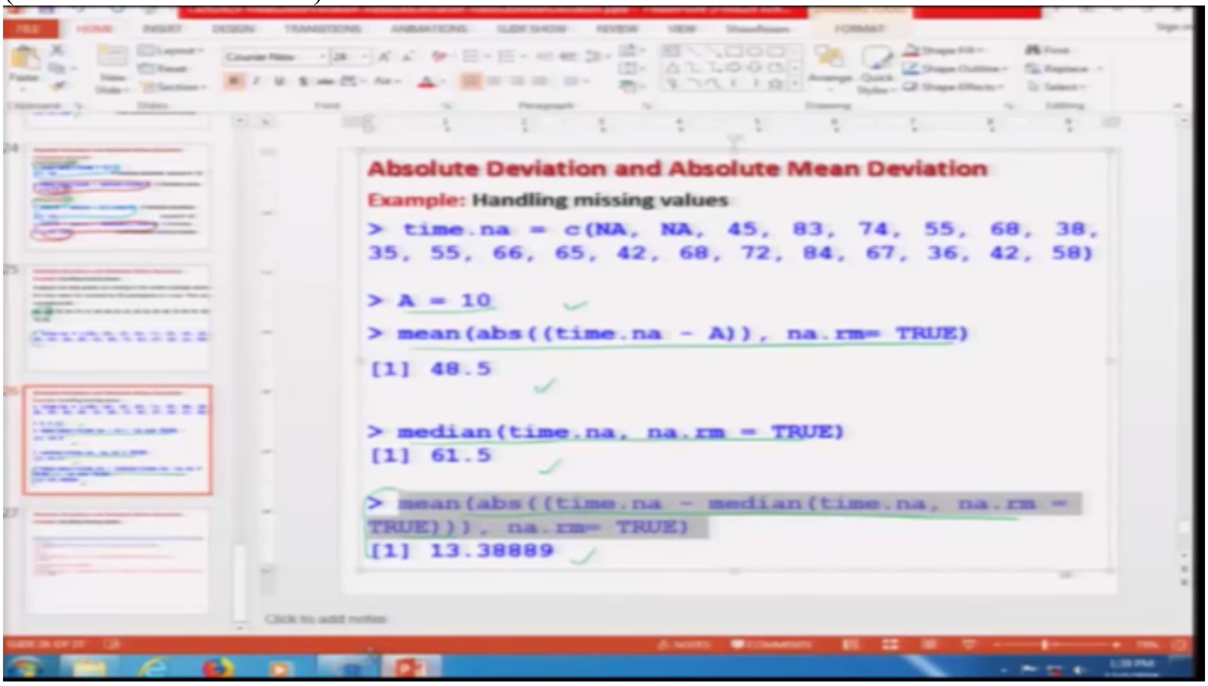
So this data has been stored in the variable time.na, and you can see here that I'm trying to choose here A = 10, and then I'm simply trying to use the command here for computing the absolute deviation when data is missing, and then I'm trying to find out the median, and then I am trying to compute the absolute mean deviation using the command that I have just discussed and these values are coming out to be here like this.

(Refer Slide Time 44:55)



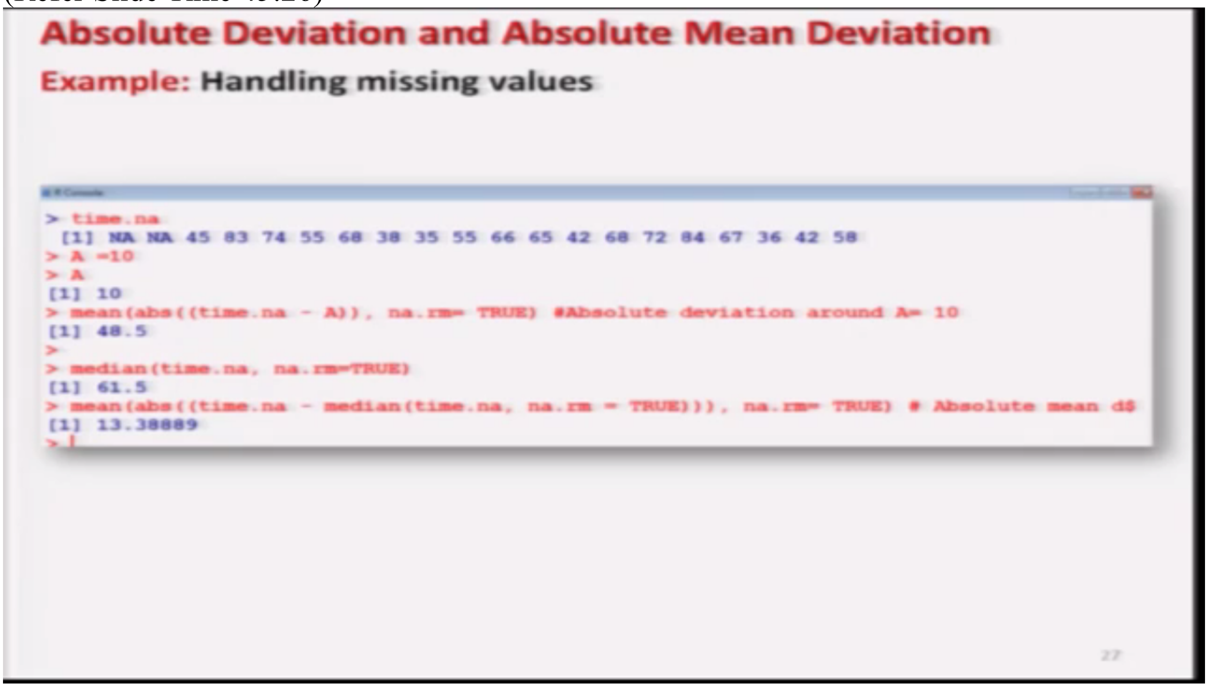
So I will try to show you on the R console also, and if you remember the time.na data has already been there. Now if I try to use the same command here, I'm getting the absolute deviation, and if I try to find out the absolute mean deviation, then it is going to be this.

(Refer Slide Time 45:19)



See here that these are the same values, the screenshot what we have just done. Right.

(Refer Slide Time 45:26)



So now I would like to stop in this lecture and you have seen that how we have formulated the measure to find out the variation in the data based on the absolute value. Notice that there is no built-in function or command in the base package of R to compute it directly. So using the earlier concepts of built-in function, we have defined the function or the command to compute the absolute deviation around any arbitrary value or say around the median that is absolute mean deviation.

You also remember, you also need to remember that the computation procedures for grouped and ungrouped data, they are also different in R Software. So you have to be careful while doing it. So I would request you that you please take some more example and try to practice it and we will see you in the next lecture. Till then good bye.

an IIT Kanpur Production