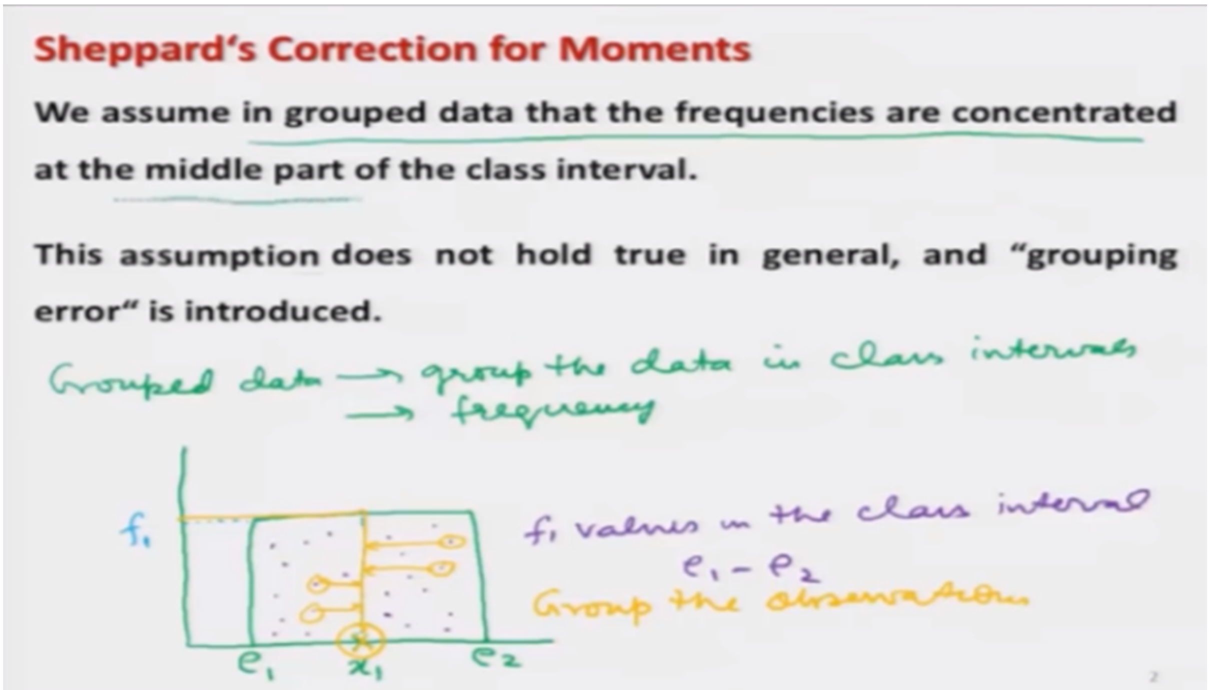
Lecture – 23

Sheppard's Correction, Absolute Moments and Computation of Moments

Welcome to the next lecture on the course, descriptive statistic with R software. You may recall that in the last lecture. We started discussion on the concept of moments, and we discussed raw moments and central moments. Now, in this lecture I will introduce you with another type of moments. Which are called as absolute moment, and I will show you, that how the raw moment, central moment and absolute moments are computed on the R software. But before we go to the concept of absolute moment, let me introduce you one small topics, which is about Sheppard correction. Okay?

Refer Slide Time (1:06)

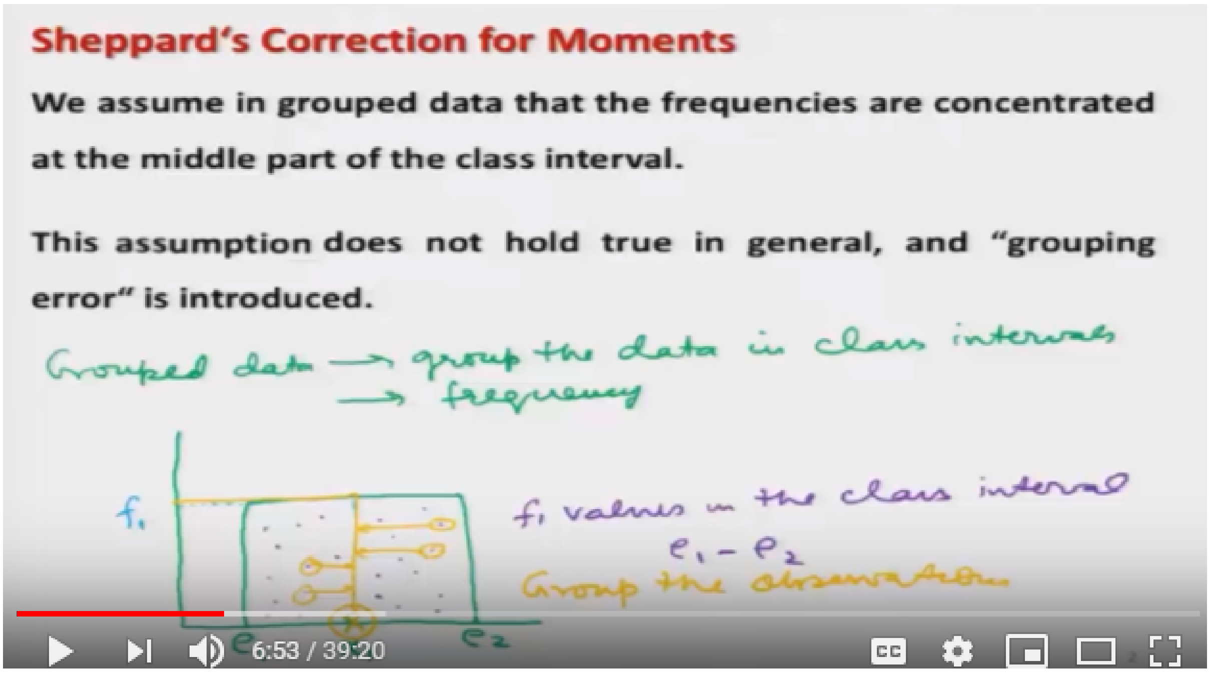


So, you may recall, that whenever we have our continuous data or say group data. What we do? We try to group them in, group the data in class intervals, and and the frequency of, that group is going to indicate, that how many values are present in that interval? Now, if you try to see what are we trying to do? We have here a sort of interval. Say here e1 to say e2. Right? and we assume, that the frequency of this interval is concentrated at the midpoint x1. if you remember x1 was the midpoint of the interval. So, we assume that on the y axis this value here, is showing here the value of see here frequency say f1. For example, first class. But now if you try to see what is happening. There were f1 values in the interval or in the class interval e1 to e2, and these values were scattered at different location inside this interval. But when you are trying to group all this information, you are assuming, that these values are concentrated at x1. So, you assume that this frequency comes here, this frequency comes here, this frequency comes here, this frequency comes here, and so on, and you assume that all these values are concentrated only in the midpoint, and this number of values is here f1. So, in some sense what are we doing? we are trying to group the observations. But when we are trying to group the observation, the information contained inside the individual observation is lost. What does this mean? That the information is lost. Suppose I have two values one is 5 and another is

10 and suppose the mid value of the interval is at 6. So, I am assuming that 5 is also becoming 6, and the value 10 that is also, becoming 6, and after this in case if you try to observe the value 6 two values of 6 like a 6 and 6. You cannot differentiate whether this value was 5 or 10, that which of the value of 6 is representing the value of 5 and which of the value of 6 is representing the value of 10 or in general we have lost the information about the individual values of xi whether the values were 5,6,7,8,9. Whatever it is we simply assume, that they are just concentrated at the middle value xi. So, you can see here when we are trying to group the observation there is some error which is introduced. and now obviously when you try to compute the moments on the basis of group data. Then this error is going to be reflected in the value of moments, and when this moment are not representing the true value, consequently they will be giving us the wrong value of mean, wrong way of variance and wrong value of other quantities which are based on moments. So, this is very important for us, that whenever we have group data, we should apply a sort of change or correction in the value of moments. So, that the modified or the corrected value of moment is used. Which in turn will give us the correct information. Okay?

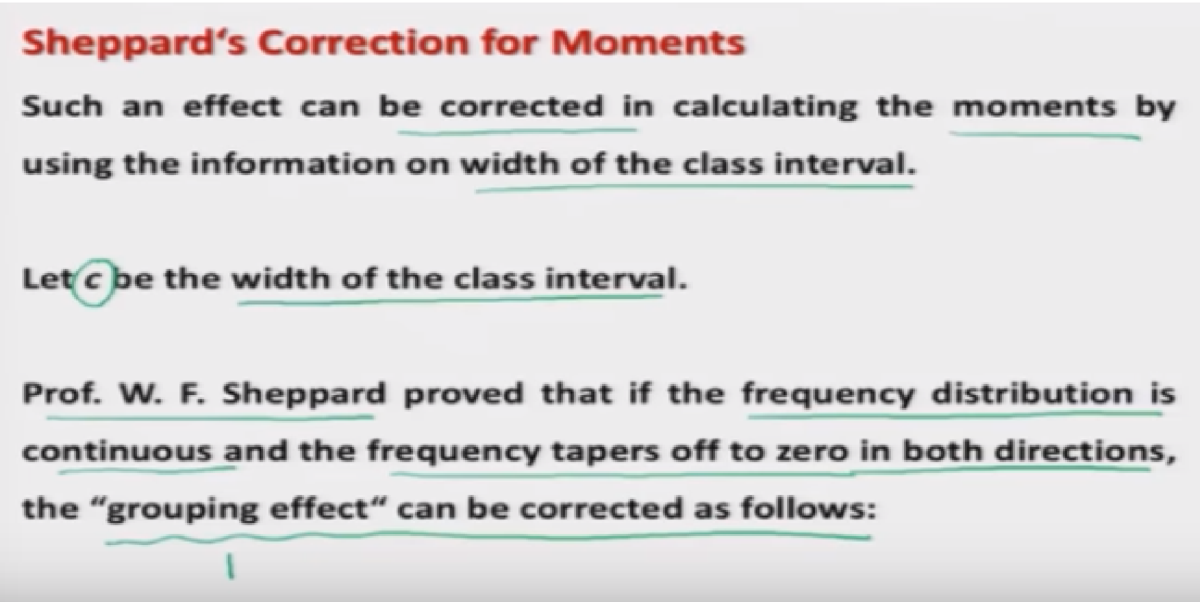
So, in this direction professor Sheppard worked, and he introduced, and he provided, some expressions, and these expressions are based only on the moment and the class interval, and he explained how this change can be made? So, that the moments are reflecting the value without end grouping error, or in simple words professor Sheppard suggested how this grouping error can be can be treated? So, let us try to start the discussion on this direction.

Refer Slide Time (6:53)



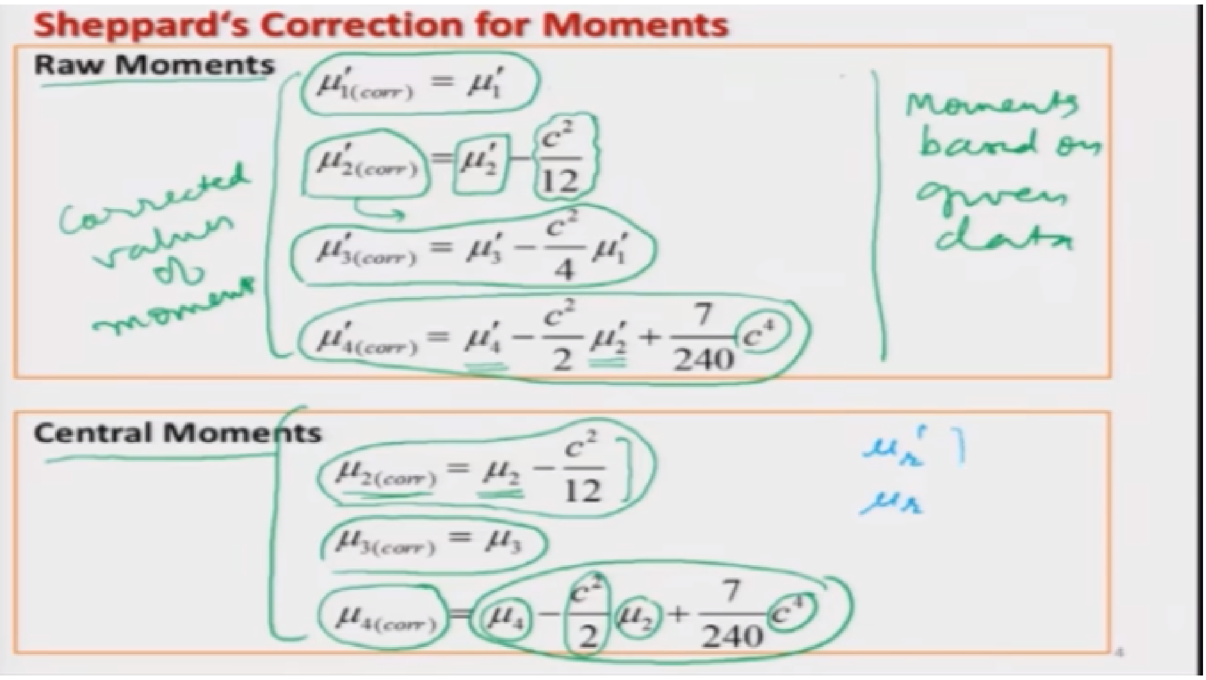
So, we assume that in a group data that the frequencies are concentrated at the middle part of the class interval, and this assumption may not always hold true and so-called the “grouping error” is introduced in the data.

Refer Slide Time (7:09)



Now, how to improve these values and how to take care of the grouping error? So, this effect can be corrected in calculating the moments by using the information on the width of the class interval. So, this is pity simple. So, let us assume that suppose small c is denoting the width of the class interval. Then professor WF Sheppard proved that if the frequency distribution is continuous and the frequency tapers off to zero in both that direction that is on the left-hand side and right-hand side, then this grouping effect can be corrected as follows.

Refer Slide Time (7:51)

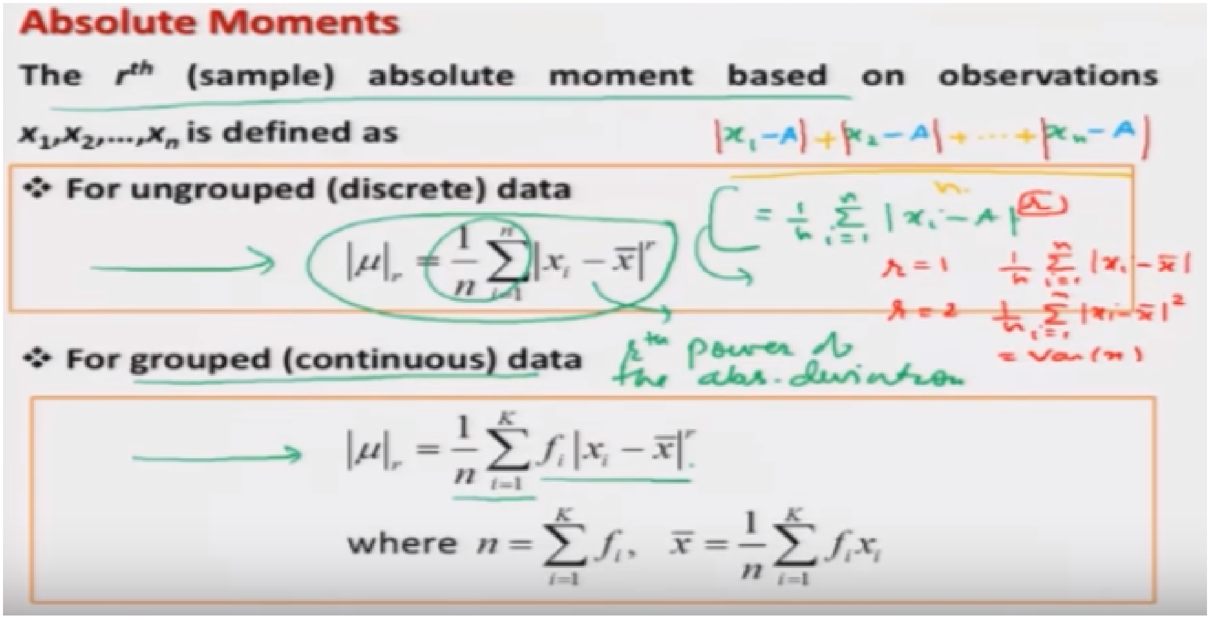


and he provided the value of raw moments and central moments after applying the changes. So, in case of raw moment Sheppard's corrections are applied as follows. On the left-hand side, I am trying to indicate the corrected values of moments, and this is the same here also in case of central moments and on the right-hand side, I am trying to indicate moments, based on given data without any correction. So, you can see here, that the first raw moment that remains the same. There is no error in this case. So, the value of the first raw moment and the so called first raw corrected moment they are the same. But in the case of second raw moment the second corrected raw moment is a function of second raw moment mu 2 prime, and it is adjusted by here a quantity c square by 12. So, what are we trying to do? That in order to take care of the grouping error, we are simply subtracting the raw moment by a quantity c square by 12. We are sees the width of the class interval, and now I will get here a new value of second raw moment in which the grouping error has been taken care and similarly in case of third raw moment the expression goes like this and the third corrected raw moment is mu 3 minus c square upon 4 multiplied by first raw moment, and similarly in the case of fourth raw moment the fourth raw moment after taking care of the grouping error is obtained by a fourth raw moment, then minus c square by 2 second raw moment plus 7 by 240 into c power of here four. We are once again I would say c is the width of the class interval, and similarly in the case of central moments also, we can modify the second third and fourth central moments. Because first central moment is always 0. So, there is no grouping error in that case.

So, thus second central moment after incorporating the grouping error or after taking care the grouping error becomes the corrected value of the second central moment is equal to the value of second central moment minus c square by 12. There is no change in the third central moment. Thus, third central moment is not affected by the grouping effect. So, the original value of the mu 3 and the corrected value of the mu 3 both remain the same. Similarly, for the fourth central moment, the value of the fourth central moment after taking care of the grouping effect is given by this. Which is the fourth raw moment mu 4 minus c square by two times. Second central moment plus 7 by 250 into serious power of here 4. So, basically these are the part which have to be taken care only during the computation, and here also you can see, that if I explain you how to compute this mu r prime. That is the raw moment and mu r which is the central moment at a central moment. Then after that at least in the first four central moment you can simply write a simple syntax in R software.

So, whatever is the expression for computing a particular moment, that value has to be adjusted just by adding and subtracting few terms as proposed by professor Sheppard. So, implementation of Sheppard
correction in R software is not difficult at all. Now, after this I will come to the aspect of absolute moments. So, how this absolute moment comes into picture?

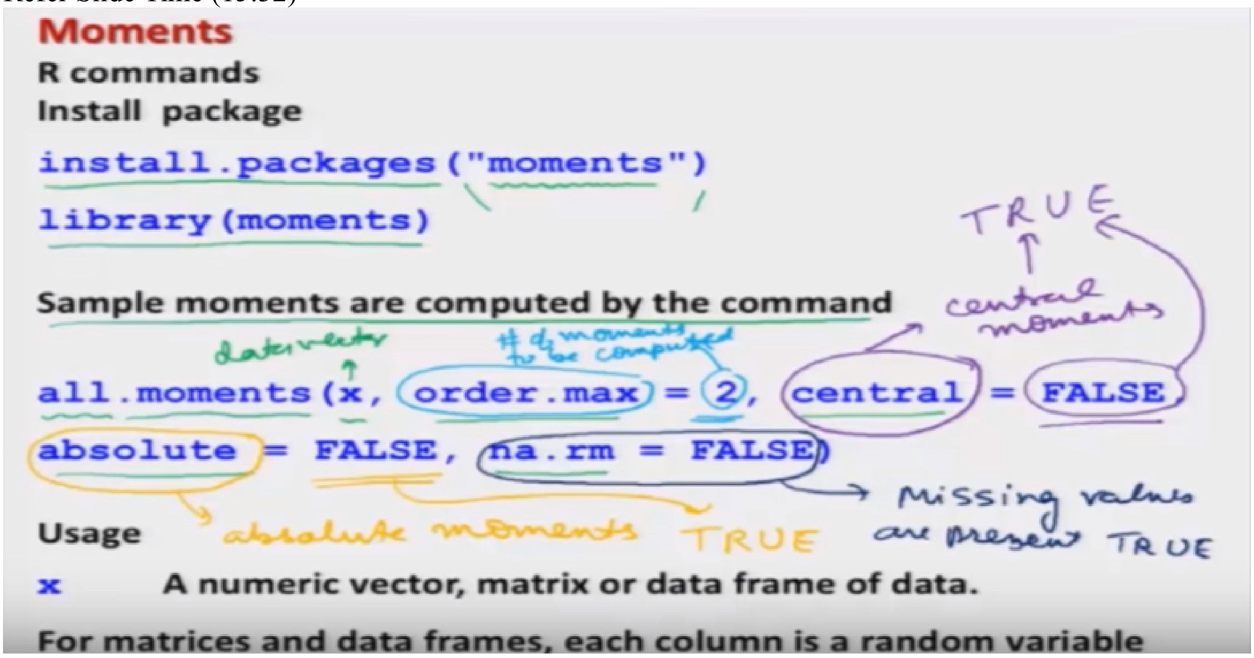
Refer Slide Time (12:24)



You have seen, that when we introduced the idea of absolute deviation what we had done? We had observations x1 x2 suppose here xn. Then, what we did? We chose an arbitrary value, and we subtracted every observation by that arbitrary value A, and after this what we did? We consider the absolute value of these deviations, and after this, we simply found the arithmetic mean of all such observations. So, this was simply 1 upon n summation I goes from 1 to n. absolute value of xi minus A. Now, in case if I try to consider here the r^{th} power of this. So, I try to add here r. So, now what will happen means if I try to take here are equal to say here 1. This will become simply a sort of absolute deviation around mean, and if I try to take here r equal to 2. This becomes I goes from 1 to n absolute value of xi minus x bar whole square. Which is same as your variance of x. So, this gives us an idea, that why not to define the r^{th} absolute moment and the quantity which you have defined here this is called as r^{th} absolute moment about A. So, the r^{th} absolute moment about arithmetic mean based on the sample observations x1, x2, xn is defined as like this. In the case of ungroup data. So, this is simply the r^{th} power of the of the absolute deviations, and after that, what are we doing? We are simply trying to find all such deviations and we are finding out its arithmetic mean. So, this is called the r^{th} absolute moment about arithmetic mean, and similarly in case of group data the r^{th} absolute moment about arithmetic mean is defined as 1 upon n summation i goes from 1 to k. f i absolutely of xi minus x bar this will power of here r.

So, you can see here that this is the same philosophy or same way of development as we have done here, the only difference is this I, have to just adjust it for the case of grouped observation, and here this x bar is going to be obtained by this expression.

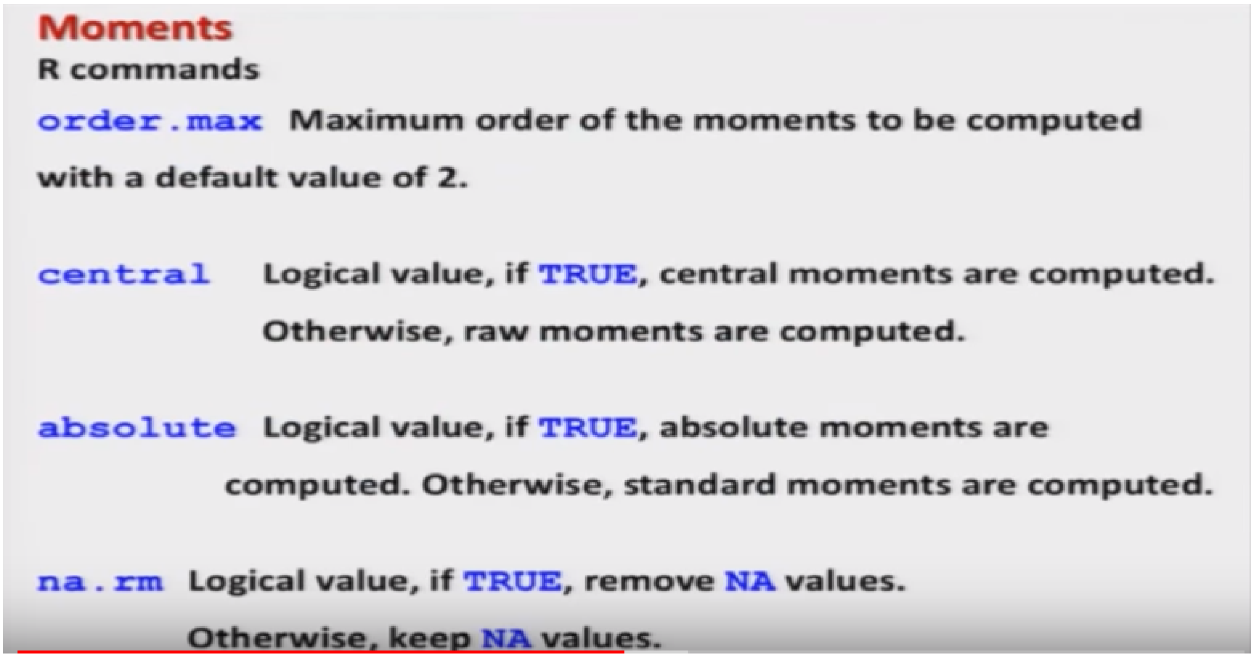
Refer Slide Time (15:32)



Now after this I, would come on the aspect that how to compute it in the R software, well when we want to compute the values of the moments on the basis of given sample of data in R software, then this part is not available in the base package of R but, in order to compute the moments, and after that I, will show you that when we are trying to measure the departure from symmetry and peakedness of the frequency curve, and we compute the coefficients of his skewness and kurtosis then we need a special package, and we need to install the package before we try to compute the moments.

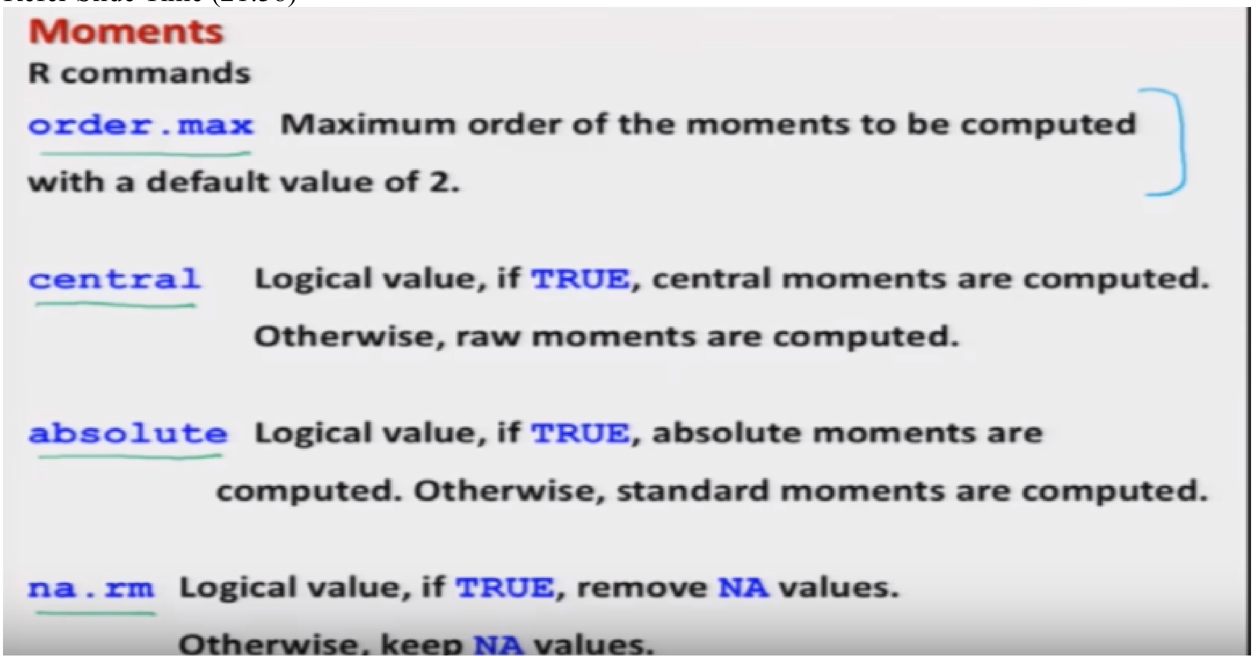
So, in order to compute the moments, we first need to install a package which is called as moments, and then we, load it as a library and then we operate. So, when we are trying to compute the moment the first step is that you try to install a package moments, and in order to do so what you have to do you have to simply write install dot packages inside the arguments within the double quotes you simply write “moments” moments and the package will be installed, and after this you need to load the package library moments in fact whenever you will need to compute the moments you need to upload this for library. Now after this, all the sample moments are computed by the command all dot moments all dot moments and this syntax or disk function has several arguments you can see here x order dot max central absolute and any dot rm just by controlling this parameter inside this argument you can generate different types of moments, raw moments, central moments and absolute moments. So, what is happening in R package in R software we have only one command all dot moments and just by giving different choices of true and false inside the argument to various parameters, we can compute different types of moments, so there is no separate command for raw moment or central moments or absolute moments. So, this is what you have to now understand which of the choice of the parameter or the values inside the argument is going to give you which type of moments Okay?

Refer Slide Time (19:27)



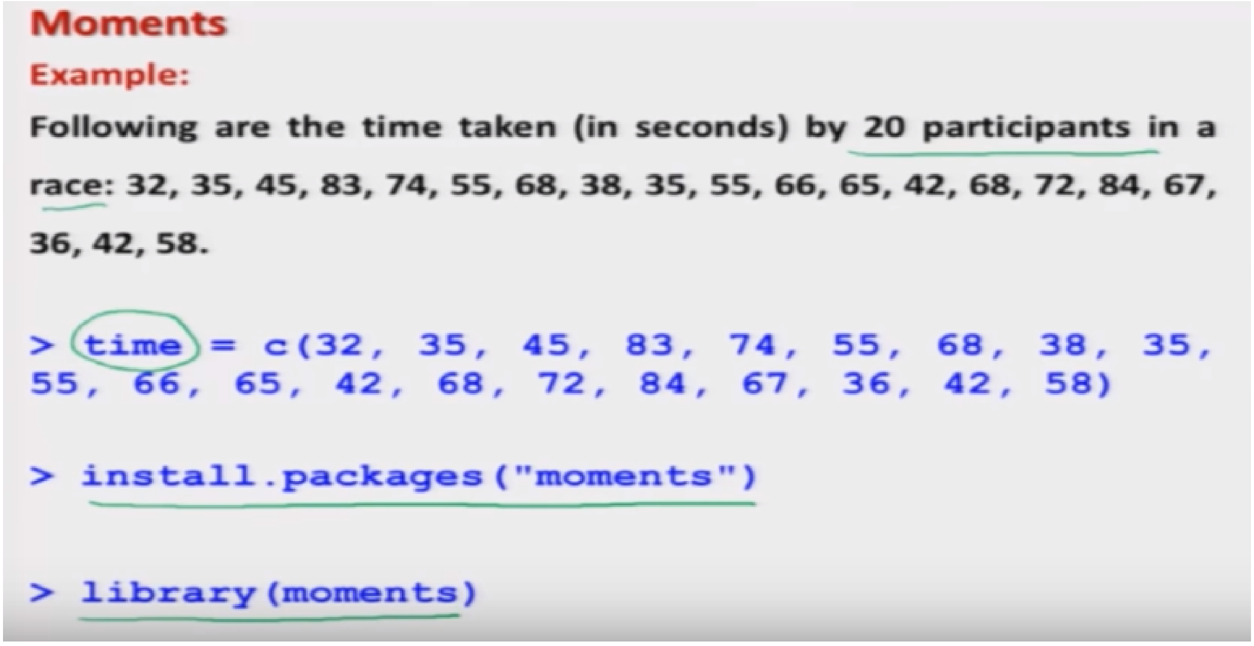
So, now firstly you let me explain you the meaning of this argument you see the first value here is x this is going to denote the data vector then there, is another parameter here all order dot max this is written here as two but this is going to give us the information on the number of moments to be computed this I explain in the next slide, here you can see here like this but I will try to explain you here also, and 2 here is that default value now in case if, you want to compute three moments for moment five moment six women you have to simply choose the appropriate value here, next command here is central, central is indicating for central moments, and the default value which is taken inside the all command is false, that is the logical false but in case if you want the central moments then what you have to do you just have to use the logical truth so in place of here false you simply try to type true and it will give you the central moment. Similarly the next option is absolute, so absolute will give you the values of absolute moments the default value which is taken in the command all that moments is false, but in case if you want to compute the absolute moment you simply have to replace this false biological true and the philosophy in tax it is known to you now this is na dot rm is equal to false, so this will try to help us in the case of missing values means if you want to compute when the missing values are present then what you need to do you simply have to change this logical false to logical to capital TRUE so, this is how just by handling different arguments with different logical true and logical false you can generate the value of different moments.

Refer Slide Time (21:38)



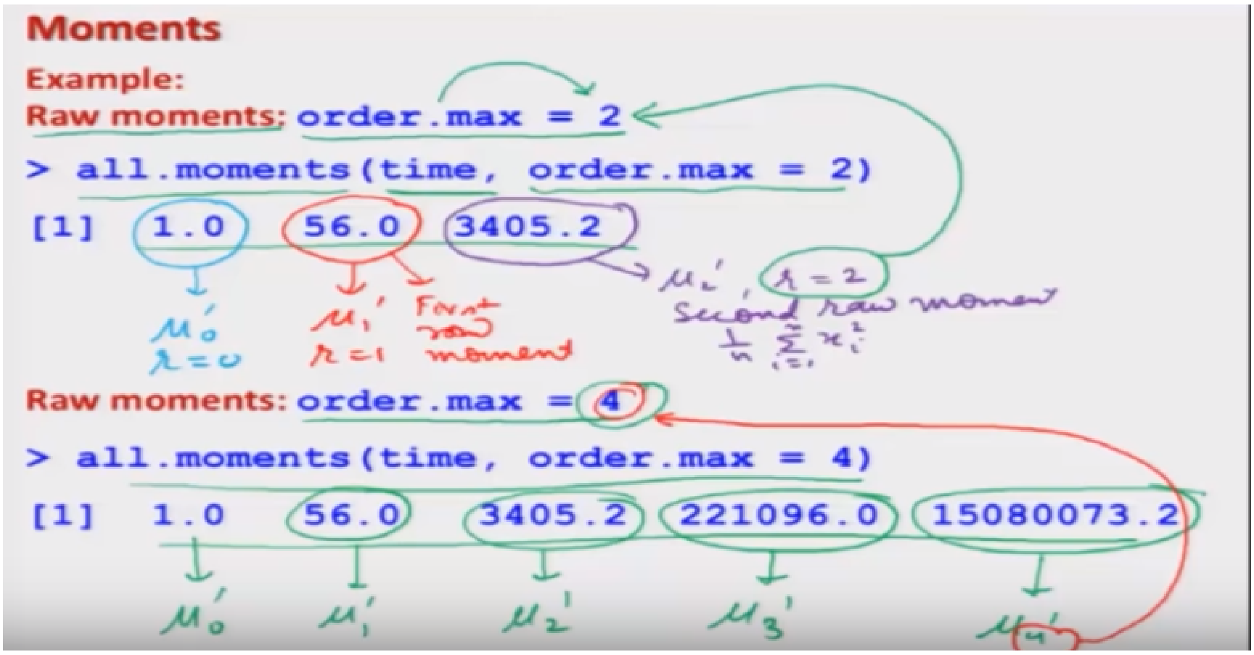
For example, in this slide I am trying to explain all these things in detail so that you can have a look so, this is about ordered max, this is about central, this is about absolute and this is about na dot rm same thing which I have just explained you Right?

Refer Slide Time (21:51)



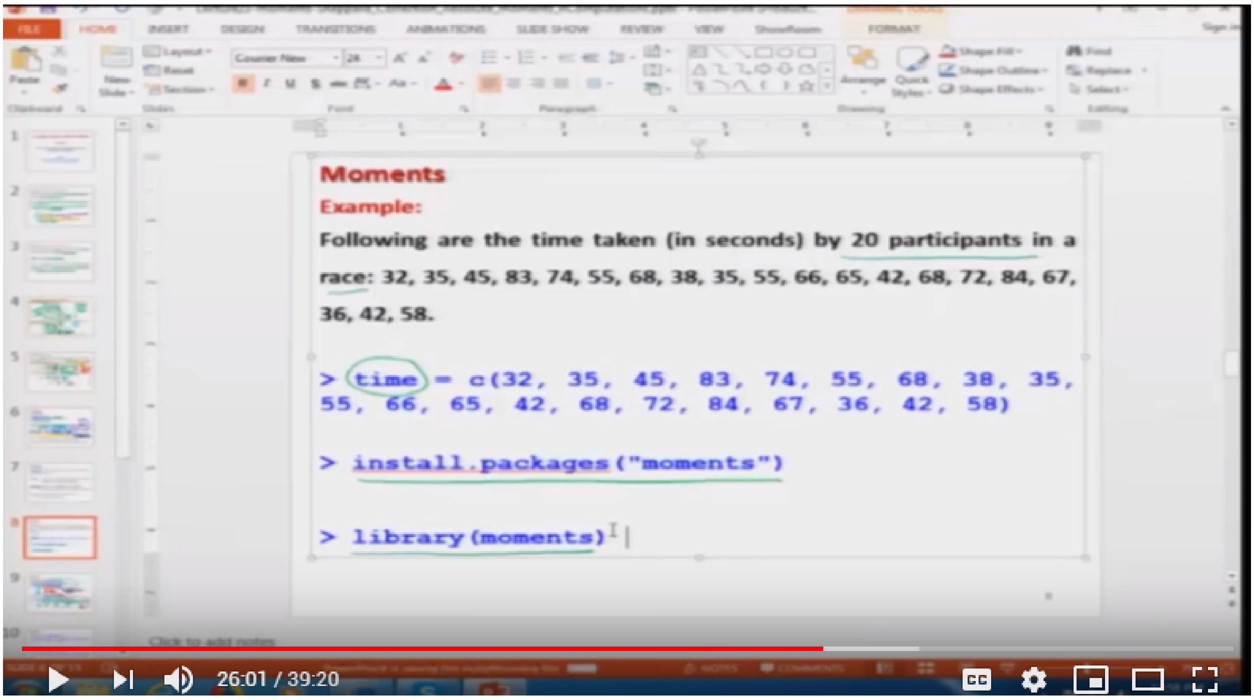
Now after this I, will try to take an example and I, will try to show you how those things are being computed so I'm taking again the same example that we have used couple of times in the earlier lectures that we have a data of 20 participants in a race, and this data has been stored in a variable here time, and now I would like to different types of moments for this data. So, as I, said first we need to use the command install dot packages to install the package moment, and then I have to load this package by using the command library moments Right?

Refer Slide Time (22:30)



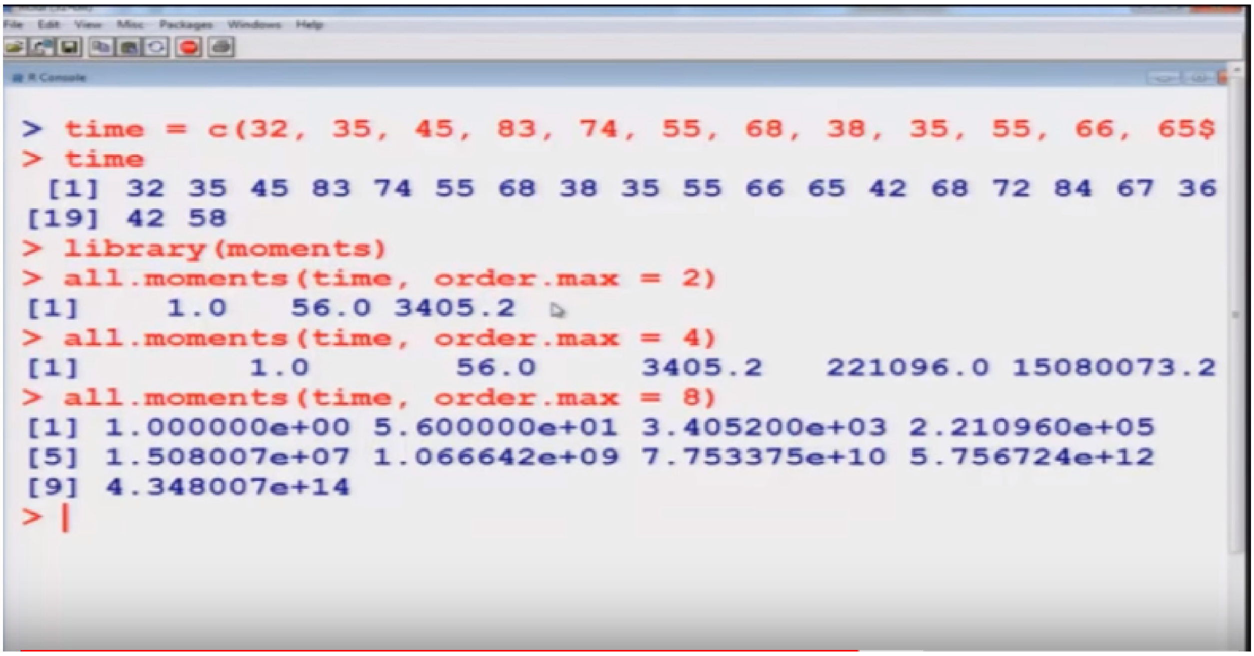
Now after this I, will show you that how to compute raw moments, how to compute central moments, and how to compute absolute moment so, first I try to take the raw of moments and suppose I, want to compute first two raw moments so, I have to control it by here order dot max 2 so, I use the command here all dot moments and the data vector time and I, have to give here order dot max equal to 2 in fact even if you don't give this option even then you will get that same outcome but my objective is to show you that how the things are being controlled. So, you can see here now once you execute it you will get here this type of outcome. Now the next question is what is the interpretation of this value add another outcomes, the first value here is 1.0 which is indeed keynoting the value of mu 0 prime that is the value of Rob moment at r equal to 0 similarly, the second value here 56.0 this is denoting the value of mu 1 prime that is the value of raw moment at r equal to 1 so, this is the first raw moment Right? This is the first raw moment and similarly the last value here 3405 point 2 this is the value of mu 2 prime that is the value of arith raw moments through moment at r equal to 2 and this is denoting the second raw moment, which is 1 upon n summation n I goes from 1 to n xi square Right? now in case if I, try to repeat the same command just by changing here the order dot max equal to four, then what will happen now you can see here that in this earlier case the maximum value of r up to which the moments are computed this is R equal to two and this is the same value which is here ordered dot max is equal to two so now suppose I, want to compute first for raw moments so, in this case I, simply have to give the value here for and then I, try to execute this command with order dot max is equal to four I get this outcome so, you can see here the first value is the value of mu 0 Prime second 56 value is the value of mu 1 prime third value 3405.2 is the value of mu 2 prime fourth value is the value of here mu 3 Prime and the last value here 15080073.2 this is the value of mu four Prime. So, you can see here this fourth which is the maximum value here for this is being indicated by this order dot max so, this will give you the first 4 moments, now before going further let me try to show you that how to compute it on the R software.

Video Start Time (26:00)



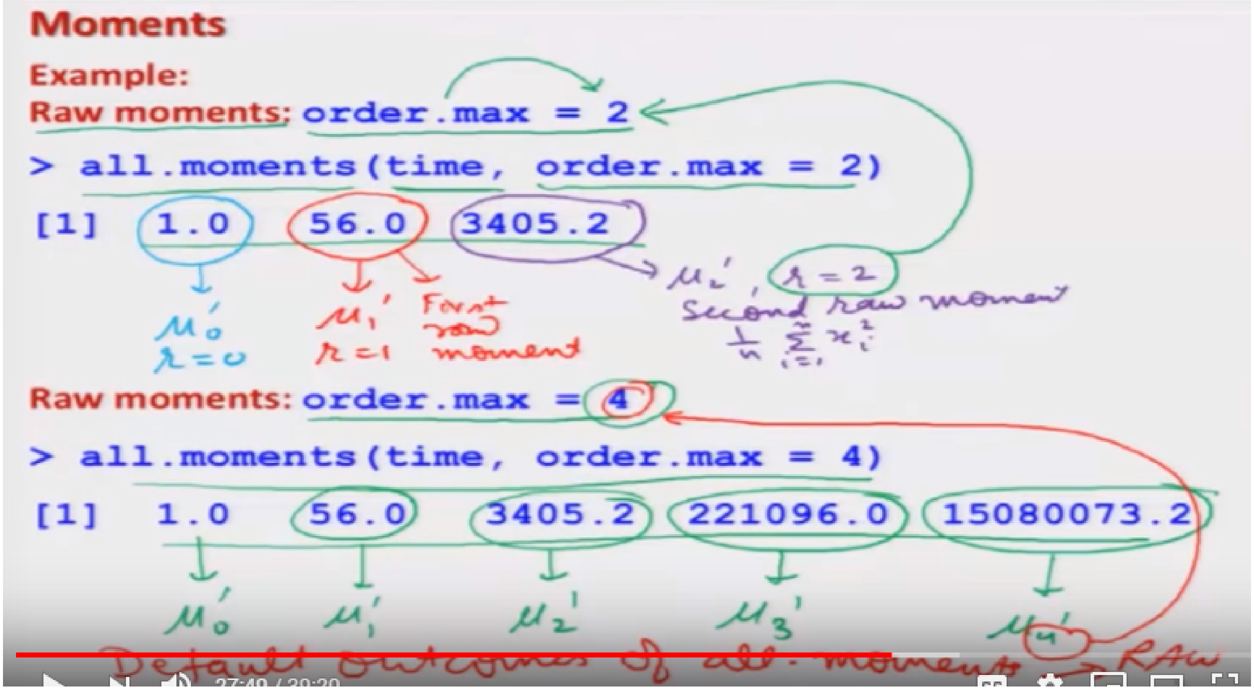
So, first I, try to create the data vector which is here like this so, you can see here this data vector here is like this and yeah I, already have installed this package on my computer but yeah you need to install the package, and I am simply trying to upload it so, now this package is uploaded moments now I, try to use the same command here all dot moments of the data vector time and I, want to compute the first to moment that is mu 0 mu 1 and mu 2 so this is like this if I try to compute say first four moments starting from mu 0 mu 1 mu 2 mu 3 and mu 4 raw moments. So, this is going to give you here like this suppose here you want to compute 8 moments you simply have to give this and you will get the outcome.

Video End Time (27:02)



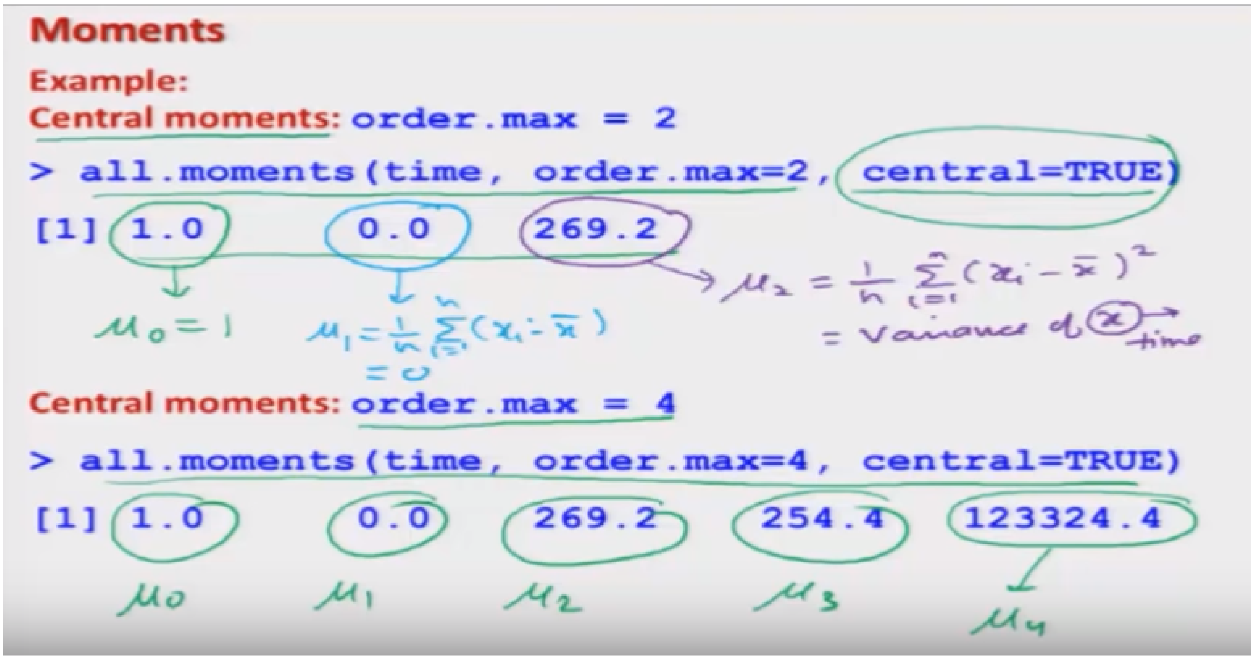
Ah very important point which you have to notice here, is that in the command all dot moments if you are not giving any option like as central or absolute or na dot rm etc., the default outcome is the raw moments, this is what you have to always keep in mind

Refer Slide Time (27:33)



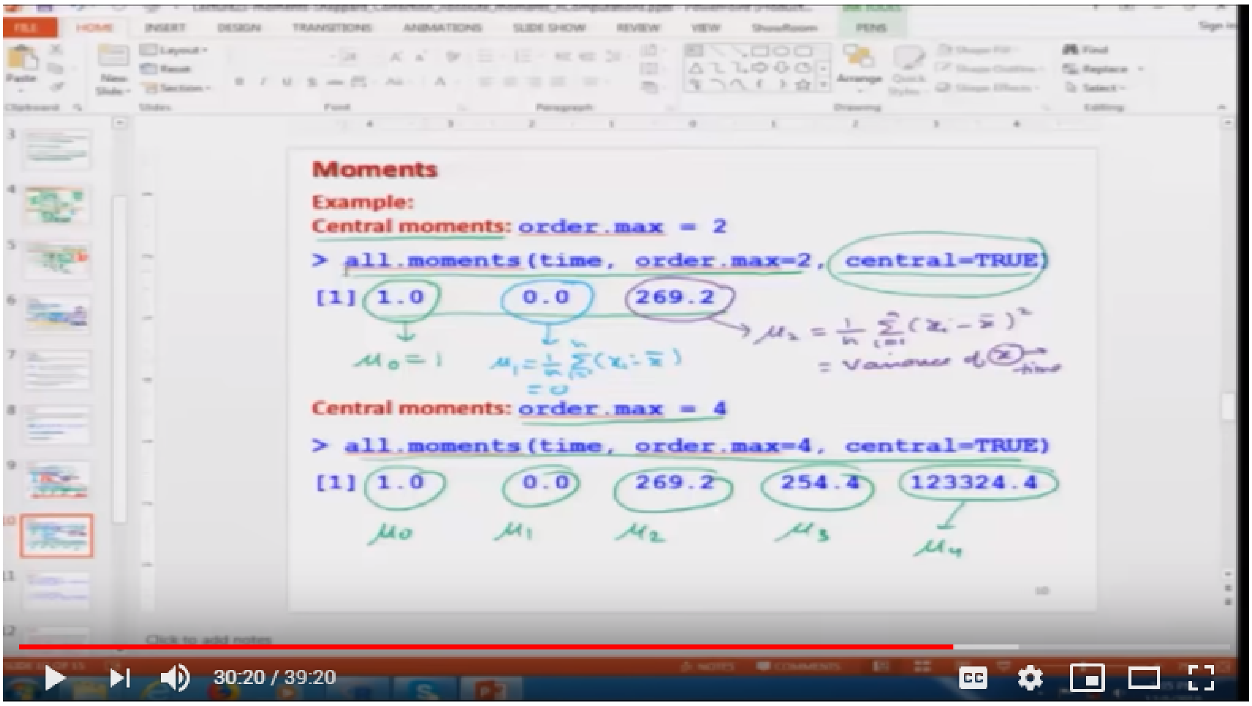
that whatever outcome we have obtained here these are the default outcomes of the command all dot moments which are the simply the raw moments Right?

Refer Slide Time (27:53)



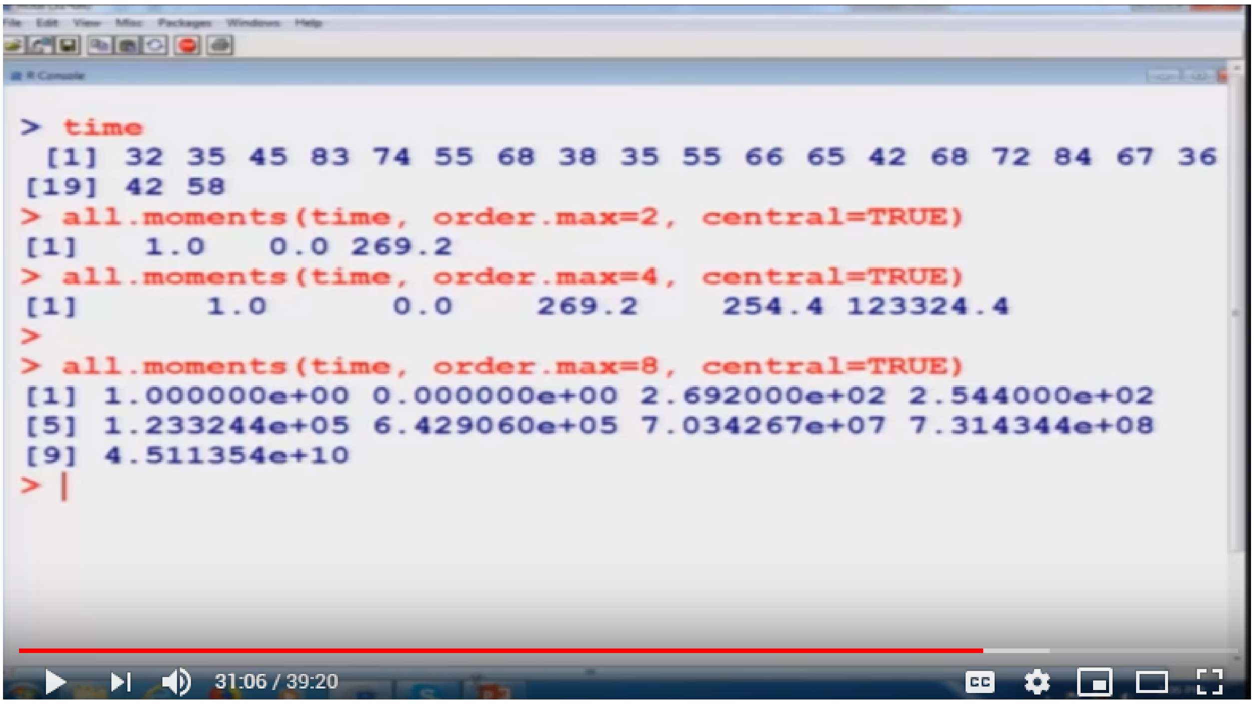
Now I, will show you how to compute the central moments so, again I, will repeat the same thing that first I, will compute the central moments up to order 2 and then up to order 4 so, what I, have to do here my command or the earlier command remains the same here all dot moments of the time data vector with order max equal to 2 what I, have to do here that I, simply have to add here one more argument central dot central is equal to logical true, that default value here is central is equal to false so, I try to adhere this central equal to true inside the argument separated by comma and the first outcome here is indicating the value of mu 0 and that was already shown that it will always take the value 1. Similarly, if you come to the second value which is 0 point 0 0.0 because this is indicating the value of mu 1 which was 1 upon n summation, I goes from 1 to n xi minus x bar and this always takes value 0. Similarly if you come to the last outcome to 69.2 this is indicating the value of mu 2 which was nothing but 1 upon n summation I goes from 1 to n xi minus x bar whole square and if, you see this is nothing but the value of variance of x or here x is actually here time the data in time vector. Now similarly if, you try to repeat the command and if you wish to compute the moments up to the order 4 then you need to make here only one chain that order dot max is equal to here four and the same come on and you can see here that this is here the outcome where first value is denoting mu 0, second value is denoting mu 1, third value is denoting mu 2 fourth value is denoting mu 3 that is that value of the third central moment and simply the last value here this is denoting the fourth central moment. So, you can see here that it is not really difficult to compute it.

Video Start Time (30:20)

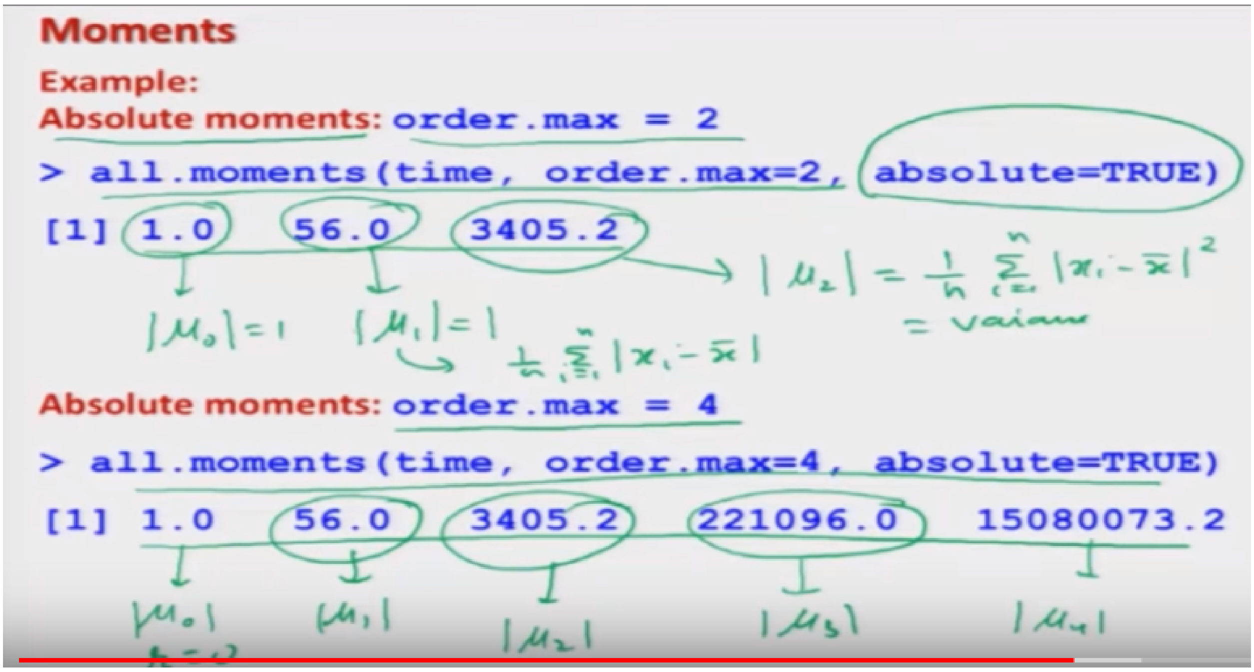


Now I, will try to show you the computation of these values on the R software on the time data, so you can see here I, try to take a like this and I, get here the moments up to order 2 that is 0 first n2o moments. Similarly if I, try to make it here 4 I get here first 4 value and similarly, if you want to make a say here first 8 moments you have to simply make order dot max equal to 8 and here is the outcome so, you can see here it is not really difficult to compute these values now let us come back to our slide and try to see finally that how to compute the absolute moments.

Video End Time (31:06)

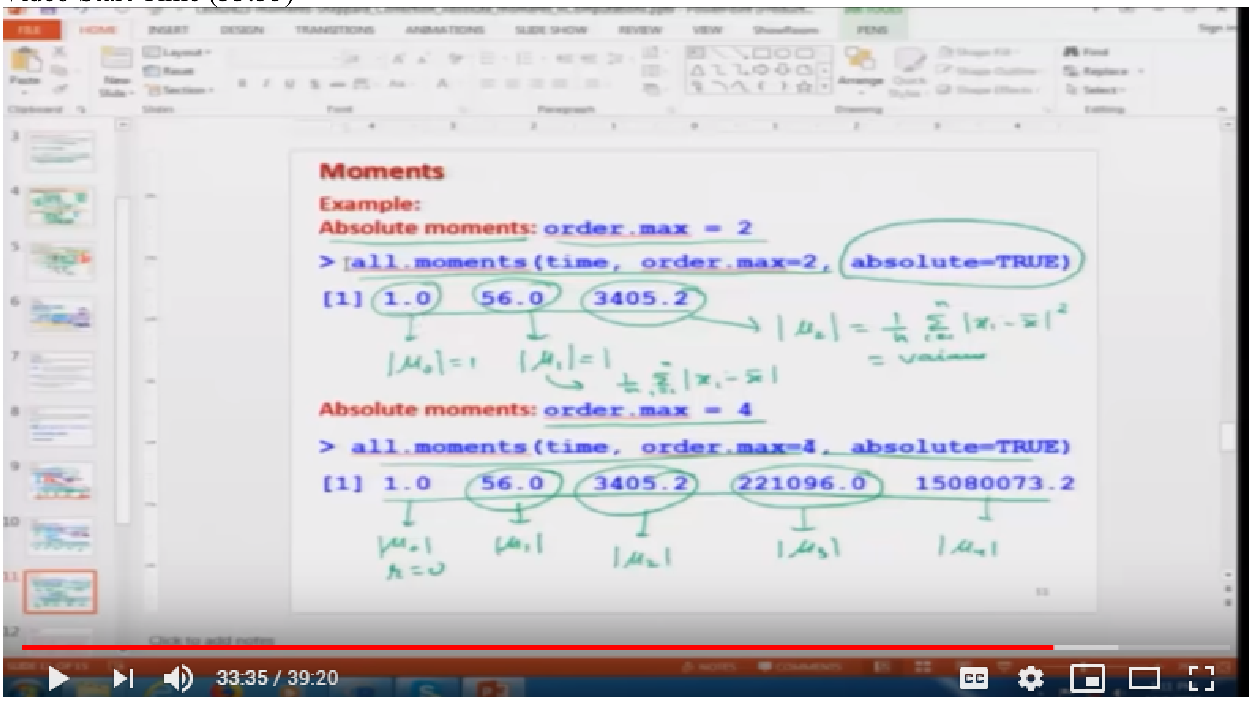


Refer Slide Time (31:09)



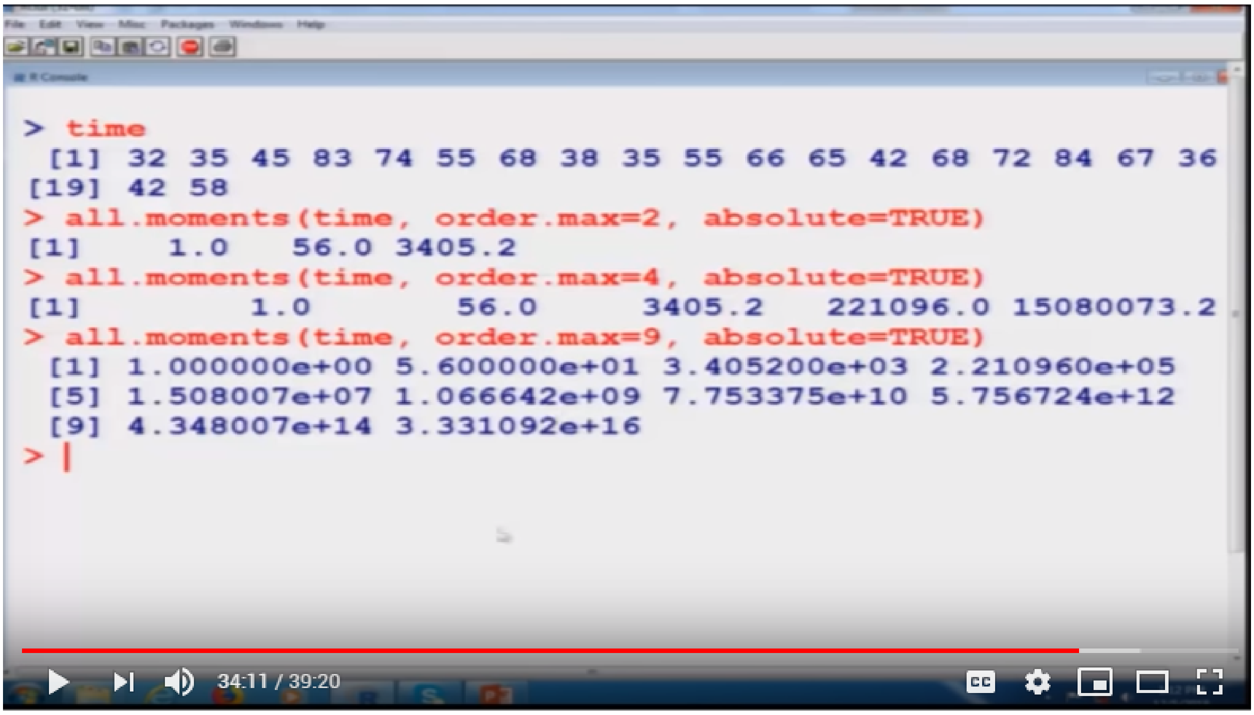
Now we have understood that as we, have computed the central moment similarly we can compute the absolute women just by controlling the argument values so in this case if I, want to compute the absolute moments say up to second order that we are equal to 0 1 & 2, my command remains the same as earlier all dot moments time order max equal to 2 and now I, add here one more option that is absolute is equal to true the default value is absolutely is equal to false but now I, need absolute moments so I'm trying to give here absolute is equal to logical true and this gives me here this these values. So, as we have done in the case of earlier example this first value that is going to give us the value of absolute value of mu 0 which is always equal to 1, and second value here this is giving us the absolute moment that is the first absolute moment, and this second value is giving us the value of second absolute moment. So, this is nothing but 1 upon n summation i goes from 1 to n absolute value of say xi minus x bar whole square which is here the variance and this mu 1 first absolute moment this is the value of 1 upon n summation i goes from 1 to here and absolute value off see here x minus x bar. Now similarly if, you want to compute the moments up to order 4 then I have to just make order dot max equal to 4 and the same command I, try to use here and this gives me this outcome so obviously this first value is going to give me the first value of absolute movement at r equal to 0 the second value is the value of first absolute movement, third value is the value of second absolute movement, and fourth value is the value of absolute movement and last value is the value of fourth absolute moment. So, this is how you can compute these absolute moment and I, will try to show you it on the R console also

Video Start Time (33:35)

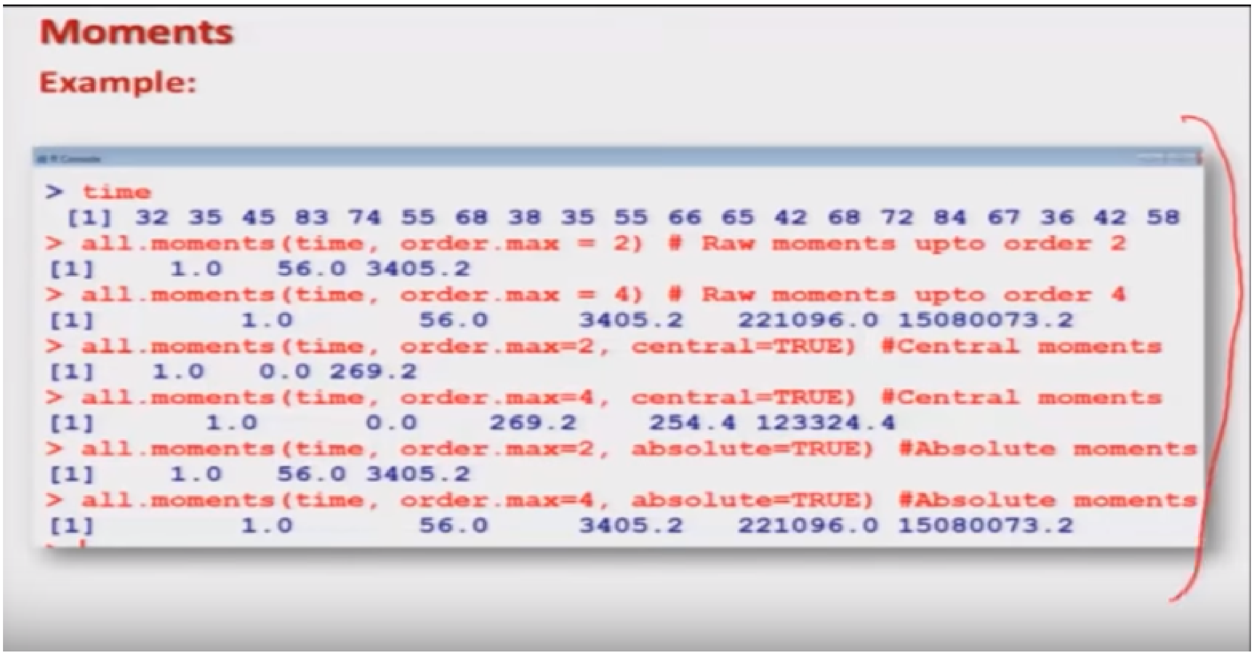


So, you can see here here were the time data, and then I, try to compute the absolute movement with order max equal to 2 which is giving me the first three values and now I, try to choose first four moments and these values are given to me like this similarly, if you want to compute any higher order value say up to nine moments, it is coming out to be like this remember one thing this counting is starting from zero so, that will so when you take order dot max equal to nine there will be ten values.

Video End Time (34:11)

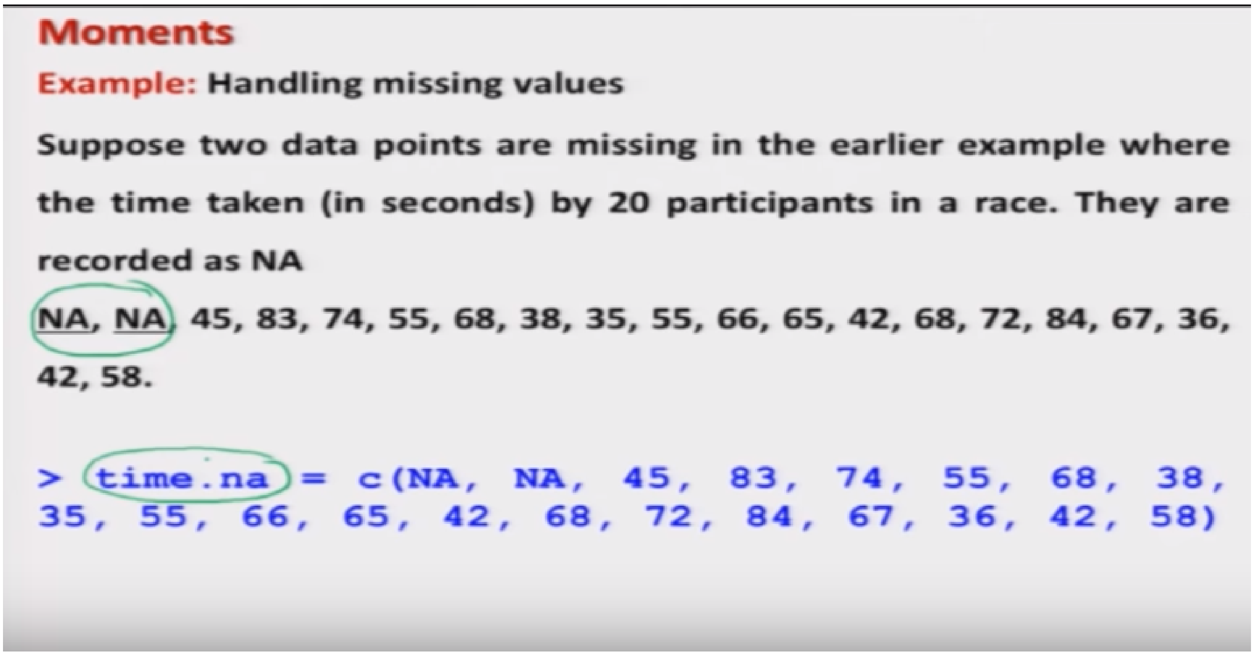


Refer Slide Time (34:13)



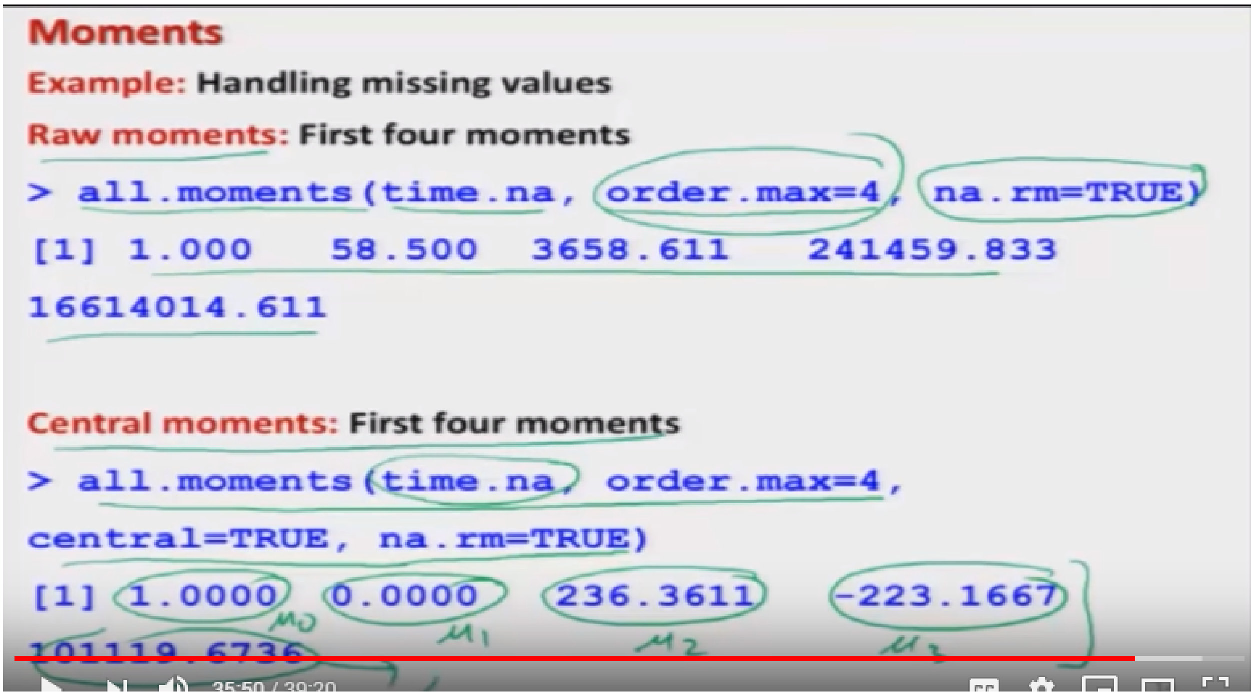
This is the screenshot what we have just done.

Refer Slide Time (34:17)



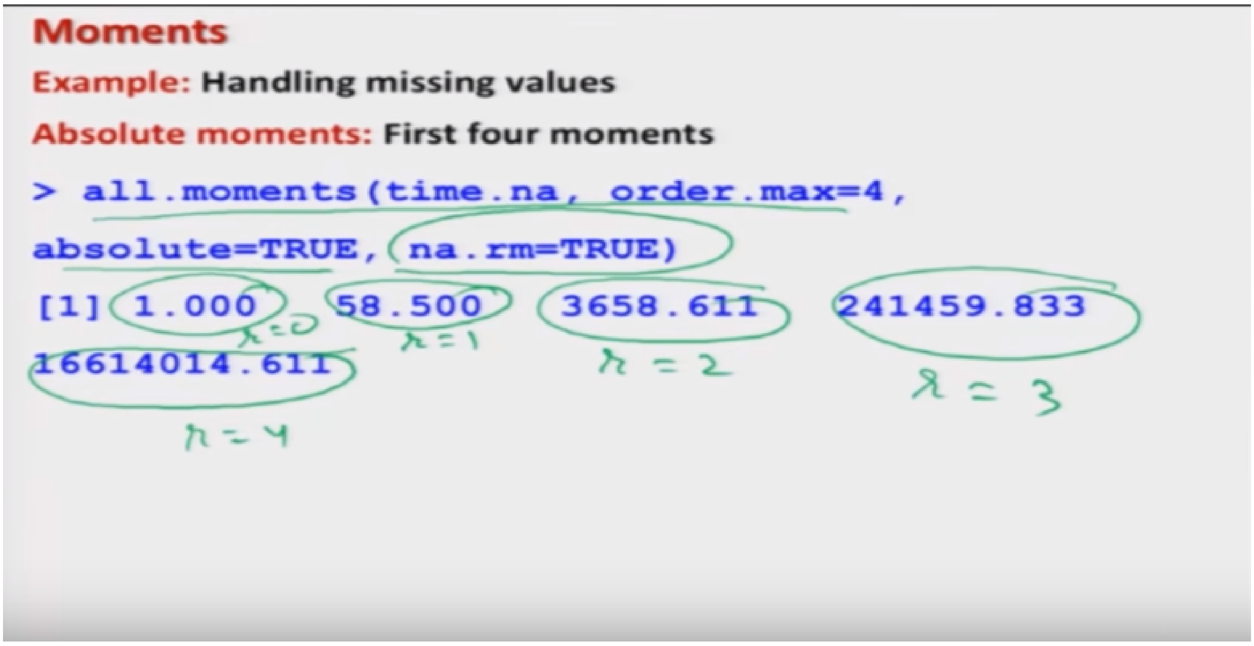
After this I, will try to show you what is the use of last option that when we have some missing value so, I will take the same data set in which the first two values have been removed and they are substituted with the na so, now and this data vector has been stored here as a time dot na.

Refer Slide Time (34:38)



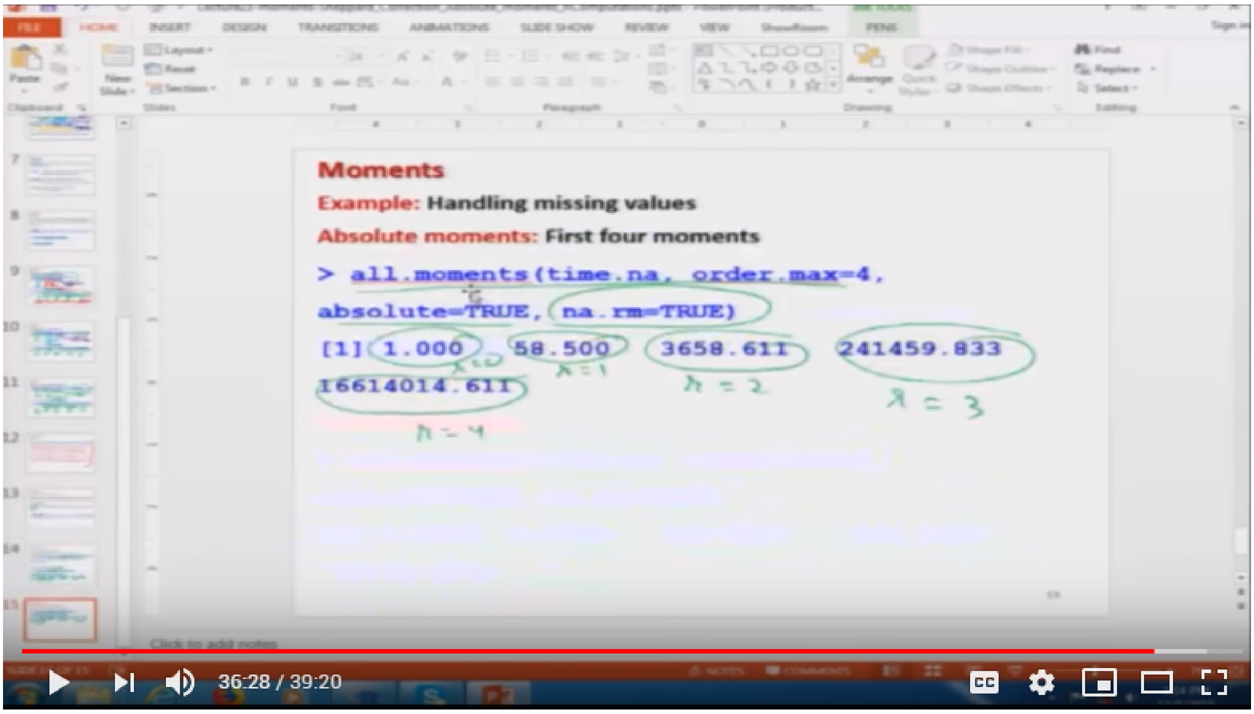
Now after this in case if I, want to compute here the rob moments in this case I, have to simply use the same command all dot moments the data vector and and suppose I, want to compute first four moments so I have to give order dot max equal to four here, and then I have to say here na dot rm is equal to true and if you don't write it then the default value here is false, and once you execute it you will get here the same outcome. So, these are the values computed in the same way as in the earlier example the only thing is this now those missing value have been removed and then the rock moments have been computed. Similarly, if you want to compute first for center moments in this case you have to use the same command that you use earlier and with the gate gate are given by time dot na and use na dot rm is equal to 2 and this will give you here the odd come and you quicker says the same outcome that did this is the value of mu zero this is the value of mu one, third value the value of mu two, and fourth values the value of mu three, and the last value is the value of mu four.

Refer Slide Time (35:55)



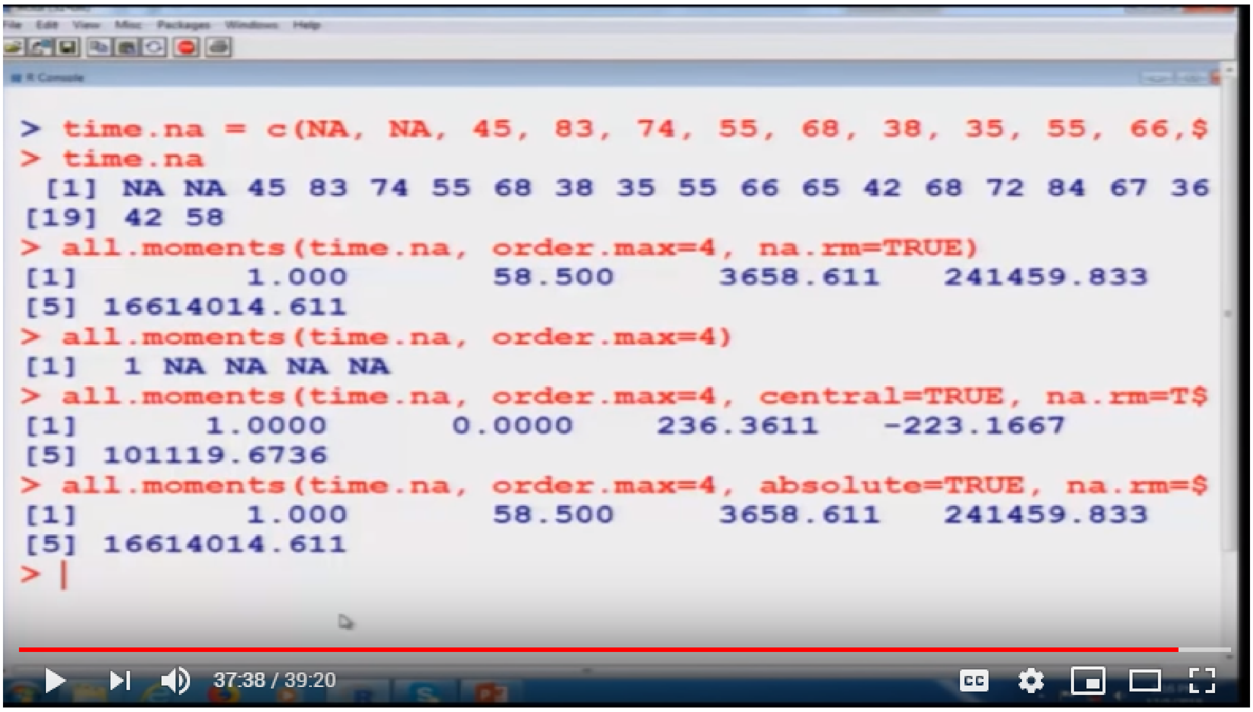
And similarly if, you want to compute the absolute moment in this case then again you have to use the same command of absolute values computation and just add here na dot rm is equal to true and you get here the value of the absolute moment for R equal to zero, the value of a salute moment for R equal to 1, that is the first absolute moment, then second absolute moment for R equal to two, the entire solute movement for R equal to three, and finally the last value for R equal to four indicating the fourth absolute moment.

Video Start Time (36:28)



Now I, will try to show you that how to get it done on the R software so, first I try to create the data vector here you can see here this is my the data, and now if I, try to compute the raw moments and if I, and suppose I show you that that if I don't use this option na dot rm then what will happen. But after execution this command on you will get the same outcome that we have used earlier now I, will try to remove this na dot rm is equal to true and you will see that it is giving you here NA NA NA NA the first value is coming out to be one because that will always remain true whatever is the value of here R and similarly if, you want to compute the central moments here just use this command and you will see the outcome here and similarly if you want to compute the absolute moments over here, then simply use the command for absolute moment and you get this outcome

Video End Time (37:38)



And similarly if, you want to compute higher order moments in the case of missing values simply try to control the value of order dot max. So now I, have given you the basic concepts of moments and I, have explained you how to compute them on the basis of given set of data. Now it's your turn try to take some datasets and try to practice and try to observe that just by choosing the logical true, and logical false inside the arguments which are given to the different parameters, how you can generate raw moments how you can generate central moments, and how you can generate absolute moments, and in the next lecture I, will show you what is the use of third and fourth order moments by considering the concepts of skewness and kurtosis. So, you practice and I, will show you in the next lecture. Till then, Good bye.