Essentials of Data Science with R Software - 2
Sampling Theory and Linear Regression Analysis
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

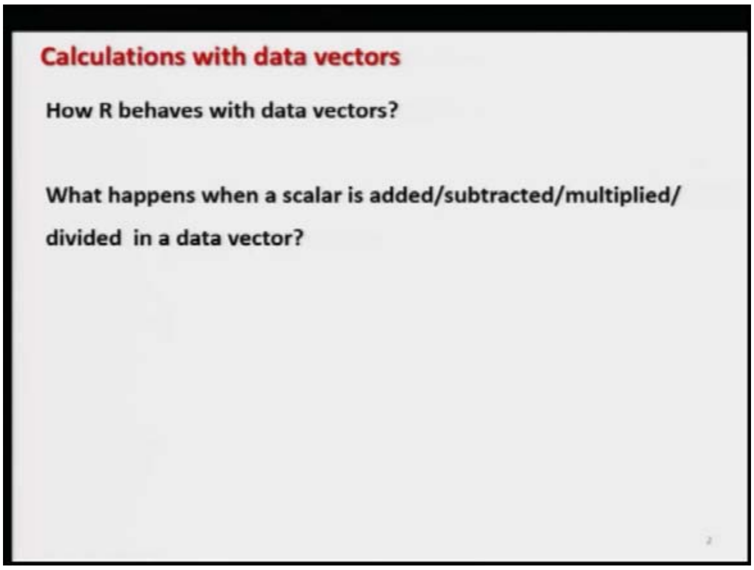
Introduction to R Software
Lecture - 04
Calculations with Data Vectors

Hello, welcome to the course Essentials of Data Science with R software part 2 where we are going to handle the topics of Sampling Theory and Regression Analysis and we are going to talk about the topics of Introduction to R Software in this module. And in the last lecture I have given you an idea that how you can make simple calculation or how you can use R like a simple calculator. Now, in this lecture I will try to give you a quick review that how R works when we have a Data Vector.

So, earlier in the lecture I have taken one value, one value at a time, but now I am trying to take more than one values in the form of a data vector and remember and recall the definition of data vector which I had discussed in the last lecture.

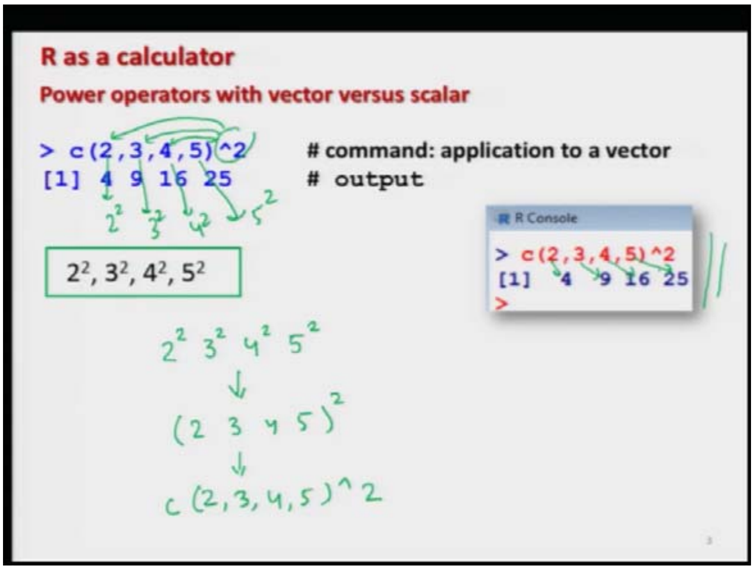
This is not going to be a vector of like matrix theory or vector calculus, this is simply a data set which has been combined using the c operator. So, when we have this type of data vector then how addition, multiplication, power operator, etc. they work in R that is what I am going to explain in this lecture.

(Refer Slide Time: 01:43)



So, my objective is very simple that I want to explore how this R behaves with data vectors, what happens when the scalar is added subtracted multiplied divided or there is a power operated in a data vector.

(Refer Slide Time: 01:55)

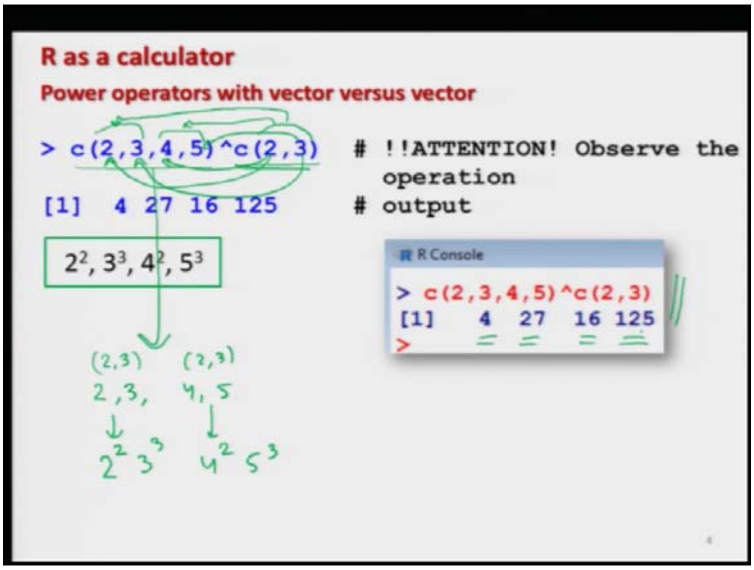


So, again I will try to take some examples and I will try to explain you how the things move forward. So, first I will try to take here the power operator, right. For example, means I will say 2 things suppose I want to find out the value of an expression which is something like 2 square, 3 square, 4 square, 5 square and so, on.

So, what I can do? That I can write down this thing here as say 2, 3, 4, 5 and I take out this square outside. So, now, in case if I want to combine it into R this 2, 3, 4, 5 can be written separated by commas with a c command and this square can be written as a hat 2 or double star 2. So, this is how the R will work.

So, the moral of the story is this once you are trying to write down a data vector with the power operator over here, then this power operator will be executed on each of this element. So, this will become 2 square, this will become 3 square, this will become 4 square and this will become here 5 square and so, you will get here an outcome you can see here 2 square is 4, 3 square is 9, 4 square is 16 and 5 square is 25. So, this power operator is distributed in all the elements in the data vector.

(Refer Slide Time: 03:18)

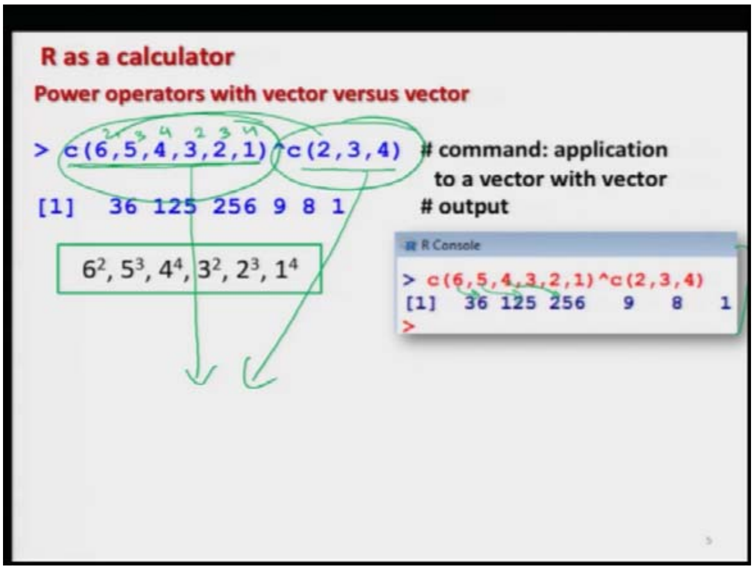


Now, the next question is that if we have a data vector like this and if this data vector is used with the power operator with a data vector with another data vector, then what happens? So, you can see here that if I try to write down here c 2, 3, 4, 5 this data vector and with the power operator hat and this is here c 2, 3.

Now, what you have to understand that this 2, 3 will come from here and it will come here this 2, 3 will come again here and it will come here. So, if I try to explain you here briefly what will happen? I have here the data vector 2, 3, 4, 5 and then on this 2, 3 will come and then once again this 2, 3 will come.

So, this will become here 2 square 3 cube and this will become here 4 square and 5 cube. So, this 2 will come over here, this 3 will come over here, similarly this 2 will come over here and 3 will come over here and you will get here an answer like if you can see here that this is 4, 27, 16, 125.

(Refer Slide Time: 04:39)

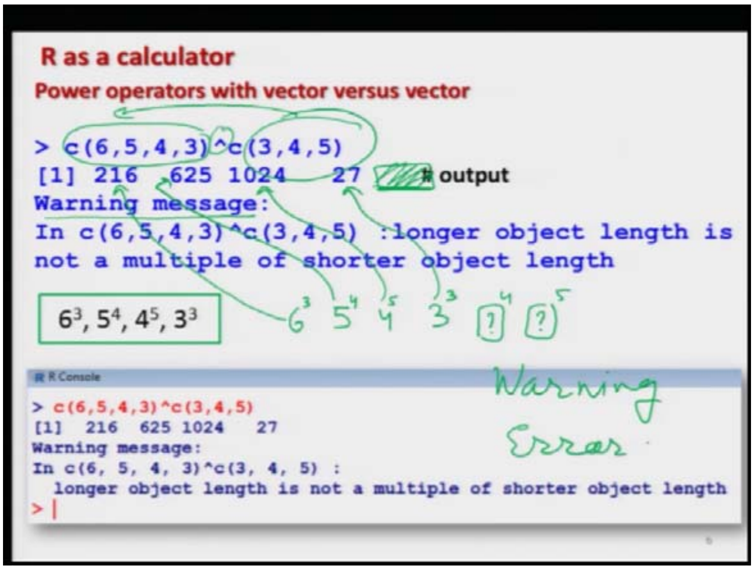


Now, if you try to see here, I am taking here another example where I have taken here 6 elements 6, 5, 4, 3, 2, 1 in the first data vector and there are 3 elements in the second data vector.

So, if you try to use here the power operator, this 2 will come over here 2, then it will be here 3, then it will be here 4 and then after this 2, 3, 4 will be once again repeated 2, 3, 4 and then you will get here an outcome like you can see here that 6 raised to the power of 2 this is 36, 5 raised to power of here cube, this is 125, 4 raised to the power of here 4 that is here 256, 4 4 is a 16, 16 4 is a 64, 64 4 is a 256 and so, on this is what you have obtained.

But you can see here one thing in this two data vectors here, you can see that the number of elements here in the second data vector they are the exact multiple of the number of elements in the first data vector. So, that is why you are getting a clear cut thing, but suppose if this does not happen then what happens?

(Refer Slide Time: 05:48)



So, for that I am trying to take here a simple example in which I have taken the 4 elements 6, 5, 4, 3 in the first data vector and the 3, 4, 5 in the second data vector now what will happen if you try to use the power operator over here? That now if I try to show you here more clearly this I have a 6 5 4 and here 3.

Now, this 3 4 5 will be coming over here with the power. So, this 3 comes over here, 4 comes over here, 5 comes over here, now once again after that 3 4 5 will be repeated. So, this will become here 3 raised to power of here cube 4 and 5, but here there is no number here this is blank.

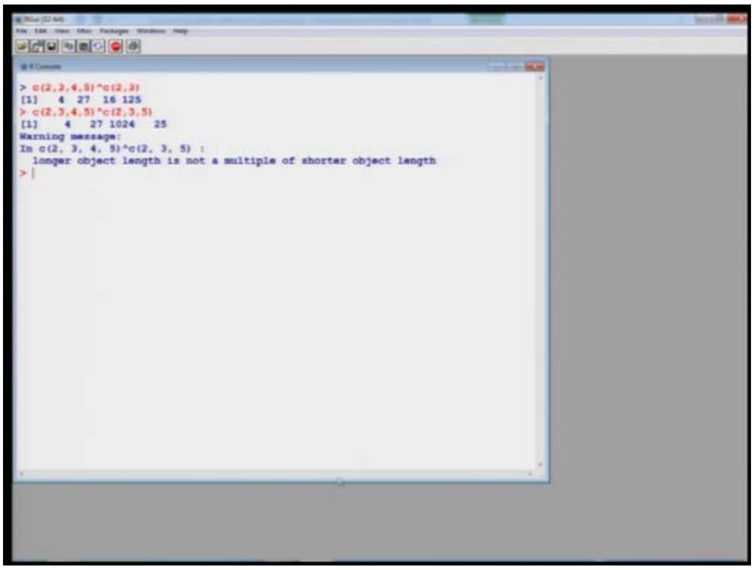
So, this will give you here a warning message, right. So, this number will be produced here you can see here this number here is 2 1 6, this 5 is the power of 4 here is 6 2 5 4 is the power of here 1 0 2 4, 3 cube is here 27. But after this there is here there is no output, right you can see here this here is blank and there is here a warning message which says that warning message in c 6 5 4 3 hat 6 3 4 5 longer object length is not a multiple of shorter objective object length.

So, this is what they mean actually and there is one thing more. Sometime you will be getting a warning message sometime you will get an getting an error message. So, there are two things warning and error. What is the difference between the warning and an error? Please try to means understand this warning and error exactly in the same way as they have their meanings. Warning is something I am warning you, but it does not mean that I will do it. But I am warning you so, but when there is an error then there is no excuse.

So, same thing happens in R, when there is an warning message, the R is trying to tell you be careful something wrong is happening, but still I am doing your job. But when there is an error message; that means, there is some mistake which is happening in the program because of which the rules of mathematics or programming are getting violated and so, R is unable to move forward.

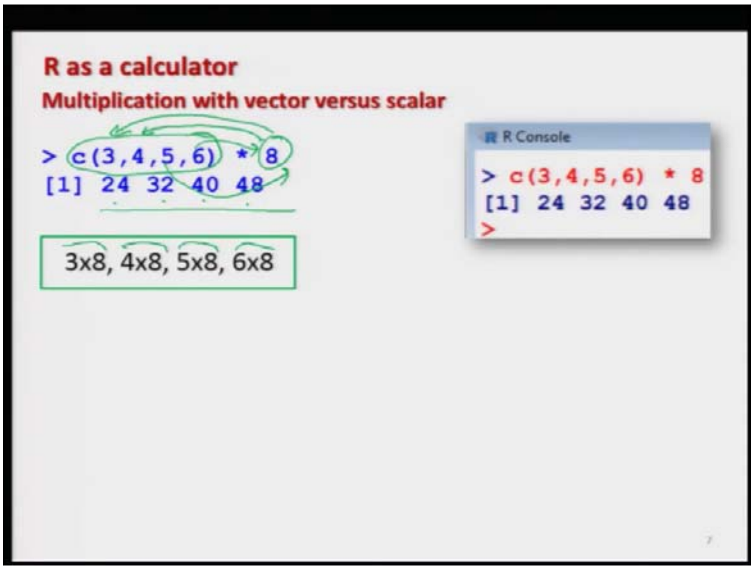
So, this is the basic difference warning and error message. So, that you get in the R console when you try to execute the program. So, let me try to show you these things on the R console.

(Refer Slide Time: 08:30)



So, if you try to see here suppose I try to take here a vector here c 2, 3, 4 and here 5 and if I try to take here c suppose here 2 comma 3 you can see here this is happening, but if I try to take it here say a number which is not a multiple, you can see here you are getting here an error message. So, the same thing you can practice you can take different types of example and you can practice there is no issue at all.

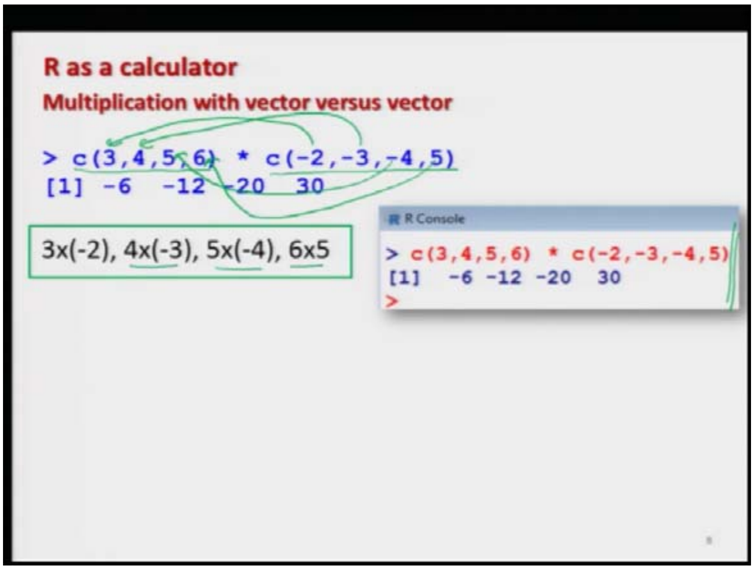
(Refer Slide Time: 09:01)



Let me take another example where I am trying to multiply a vector by a scalar, I mean there is only one number. So, if I try to take here a data vector 3, 4, 5, 6 but I simply try to multiply it by here 8. Then you can see here the outcome here is 24, 32, 48; that means, R has a peculiar characteristic that this number 8 is being operated on all the elements.

So, this is happening that 3 is being multiplied by 8, 4 is multiplying by 8, 5 by multiply by 8 and 6 is multiplied by 8. So, basically this is the outcome here 24, 32, 40, 48 which is this outcome 3 into 8, 4 into 8, 5 into 8, 6 into 8. So, remember one thing this is a very peculiar characteristic of R that whatever operation you are going to do here that is being operated on each and every element of the data vector.

(Refer Slide Time: 09:57)

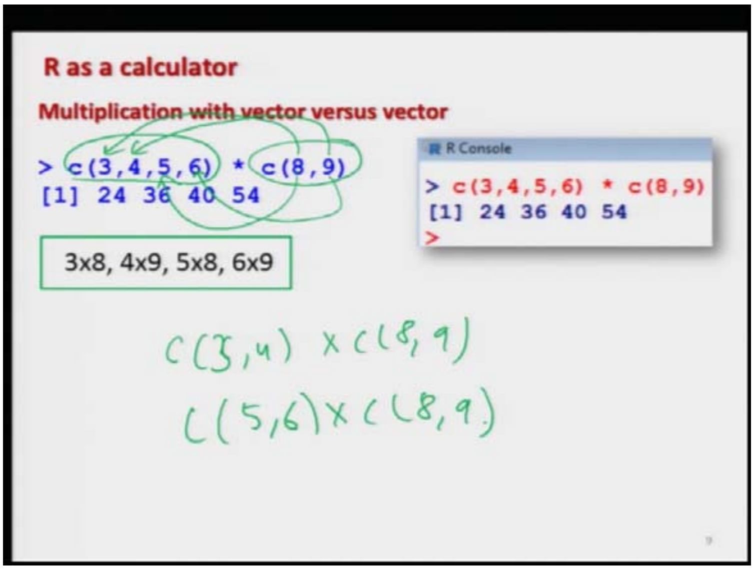


And similarly, if you try to extend this example and if you try to multiply 2 data vectors. So, if I try to take it here a data vector c 3, 4, 5, 6 and if I try to take here means another data vector minus 2, minus 3, minus 4, and 5.

Then what will happen? The same rule what we have understood that will be followed here that first element will be multiplied with the first element, second element will be multiplied with the second element, third element will be multiplied with the third element and fourth element will be multiplied with the respective fourth element.

So, you can see here this will be essentially be the element wise multiplication, 3 into minus 2, 4 into minus 3, 5 into minus 4 and 6 into 5 this will be your here the outcome exactly in the same way, but here you can notice look at the number of elements in the two vectors are exact multiples.

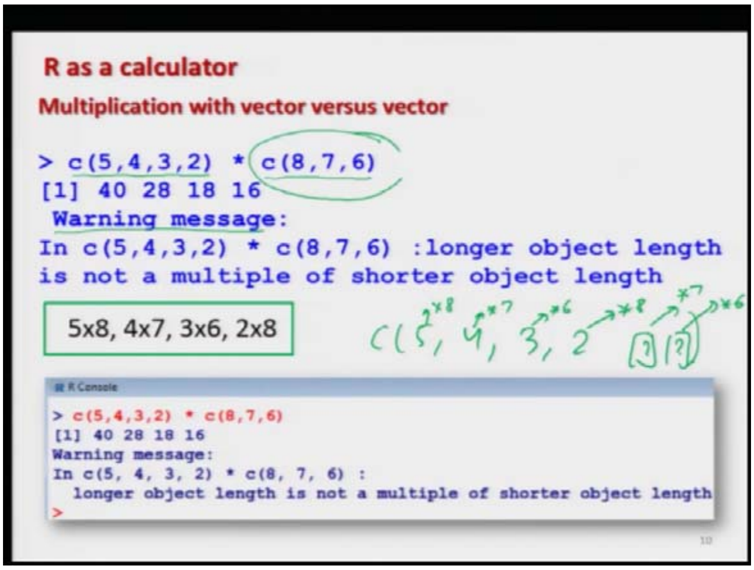
(Refer Slide Time: 10:53)



Here in this case the exact multiple is 1, but there can be a different case like this one where I am trying to take here one data vector as say here c 3, 4, 5, 6 and say another data vector here is with only two elements c 8, 9. So, in this case what will happen? That means again this multiplication will be operated on the entire element and it will be something like this c 3, 4 into c 8, 9 and then c 5, 6 into c 8, 9 right.

So, this 8 and 9, 8 will be multiplied here, 9 will be multiplied here, and similarly 8 will be multiplied here and 9 will be multiplied here. So, you can see here this will become 3 into 8, 4 into 9, 5 into 8, 6 into 9. So, this distribution is getting implemented over here in an element wise way, right.

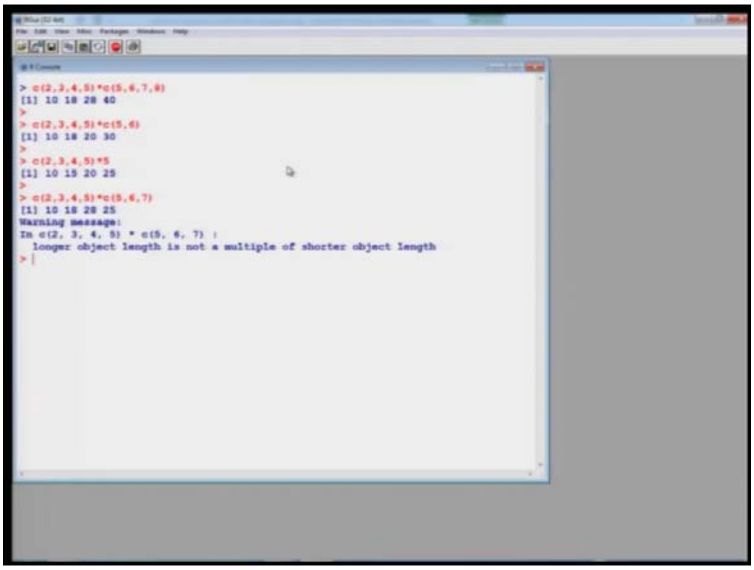
(Refer Slide Time: 11:52)



And similarly, if you do not have the exact multiple of the numbers in the two vectors, then you will have a warning message. For example, if I try to take here 4 elements in the first data vector and 3 elements in the second data vector which are not exactly multiple. So, if I try to multiply c say here c 5, 4, 3 and here 2 right.

So, now it has been multiplied by here c 8, 7, 6, this will be multiplied by 8, this will be multiplied by 7, this will be multiplied by 6. Now, once again 8 7 6 will be repeated, so this 2 will be multiplied by here 8, but after that what to multiply with 7 and 6, this is missing. So, that is why you are getting here a warning message. So, this is how multiplication operator works in the case of R software.

(Refer Slide Time: 12:56)

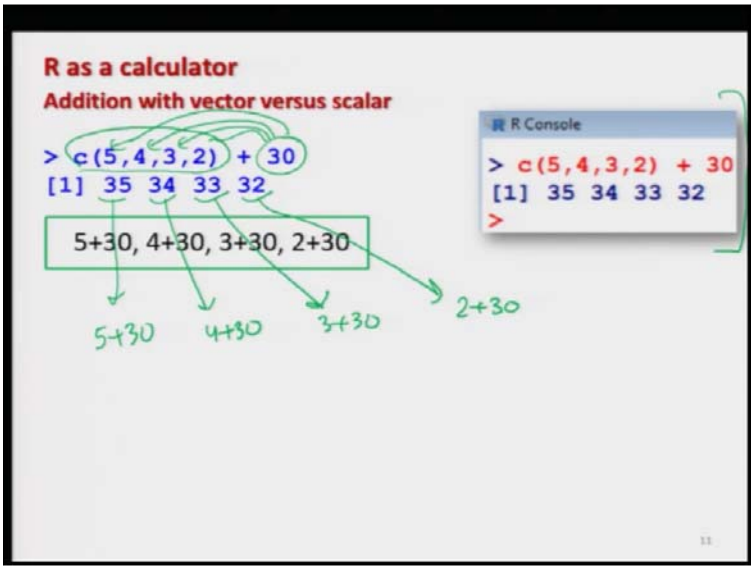


So, let me try to show you here in this R console also. For example, if you try to take here a data vector c say 2, 3, 4 and here 5 and if I try to take it here c say here c 5, 6, 7, 8. So, you can see here the number will be say 2 multiplied by 5 10, 3 multiplied by 6 18, 4 multiplied by 7 28 and 5 multiplied by 8, this is 40.

And similarly if you try to take here say here only two elements c 2, 3, 4, 5 multiplied by 5, 6. So, you can see here the outcome is 2 multiplied by 5 10, 3 multiplied by 6 which is 18, 4 multiplied by 5 this is 20 and 5 multiplied by 6 which is 30 right. So, they are getting distributed. Now, I try to take here only here one say this number say here is 5. So, you can see here this 5 is multiplying all the elements in the data vector. So, this is 2 into 5 which is 10, 3 into 5 which is 15, 4 into 5 which is 20 and 5 into 5 which is 25, right.

So, this is how the things will happen. Now, in case if I try to take an equal number of elements for example, if I try to take 4 elements in the first data vector and 3 elements in the second data vector you can see here, this will give you the answer on the same rule, but with a warning message because the last two elements in the second data vectors which are 6 and 7 are not finding any place or any number in the first data vector where they can multiply the numbers, right. So, that is pretty simple and stayed forward, right.

(Refer Slide Time: 14:53)

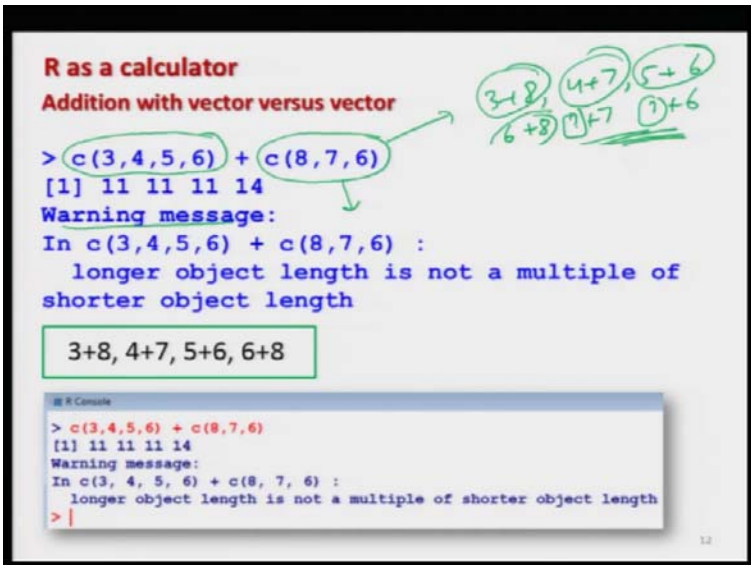


Now, I move forward and if I try to show you the addition and this subtraction on this data vector, they have the same rule that if you try to take a vector here. Suppose I take a vector here 6, 5, 4, 3, 2 and then I try to add it with here a scalar quantity say here 30.

So, now in this case what you can see? This number is this number 30 it is getting added in every element of this vector and essentially you would get here an outcome which is here 35 which is say here 5 plus 30. Then you are getting here an outcome 34 which is 4 plus 30 and then you are getting a 33 which is here 3 plus 30.

And this here number here 32 which is here 2 plus 30, right, and this is here the screenshot of the outcome. So, you can see here the rules are the same that addition operator is being distributed over the entire data vector and every element is getting added with the number 30.

(Refer Slide Time: 15:59)



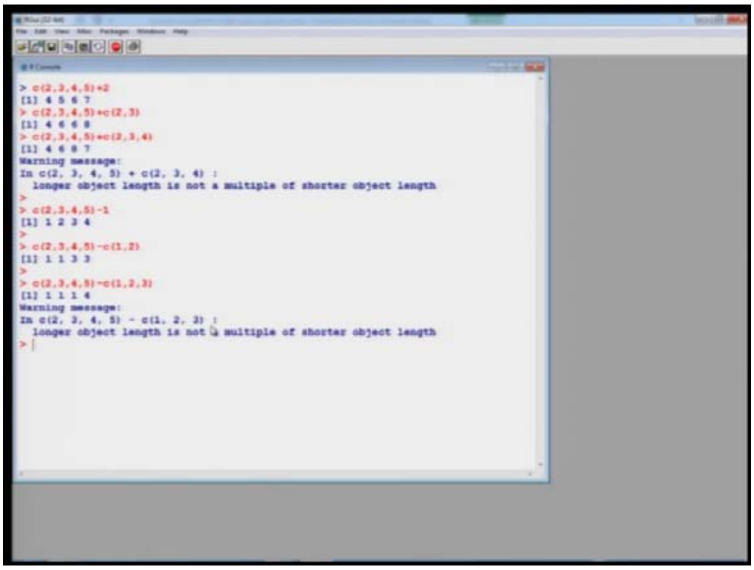
And now if you try to use the same logic means I can add and subtract here 2 data vectors also for example, if I try to I have just taken an example to show you because now you can understand what is happening. Suppose, I try to take here a data vector 3, 4, 5, 6 and second data vector which has got here say 3 number which are not really the exact multiple of the numbers in the first data vector, then again it is giving me an answer here 3 plus 8, 4 plus 7, 5 plus 6.

Now, after this 8, 7 and 6 are to be added to some number. So, in the first place it is finding here 6, but after that it is not finding here any number where it can add. So, it is giving me an answer here 8 3, 11 7 4, 11, 5 6, 11 6 and 8 here 14, but after that there is no number.

So, that is why it is giving me a warning message. But if there are only say two numbers in the second data vector which is exact multiple of the number of data points in the first vector, then this error message will not come this warning message will not come. And this I will try to take here some example and I will try to show you.

And you know once you have understood the addition the same thing happens with the subtraction also. So, instead of using my slides I will try to take here several examples to explain you what is really happening once you are trying to do it in the R console with this addition and subtraction, that is pretty simple.

(Refer Slide Time: 17:58)



So, if I try to take here a data vector c 2, 3,4 comma 5 and suppose if I try to add it here with some scalar say here 2, then it is 4, 5, 6, 7 which means 2 is being added to every element in the data vector which is 2 plus 2, 3 plus 2, 4 plus 2 and 5 plus 2.

Similarly, if I try to take here a data vector instead of a scalar say 2 comma 3, then you will see here that first 2 elements 2 and 3 are being added with 2 and 3 and next 2 elements here 4 and 5 they are also being added with 2 and 3. So, 4 plus 2 is here 6 and 5 plus 2 and 5 plus 3 here is 8 right.

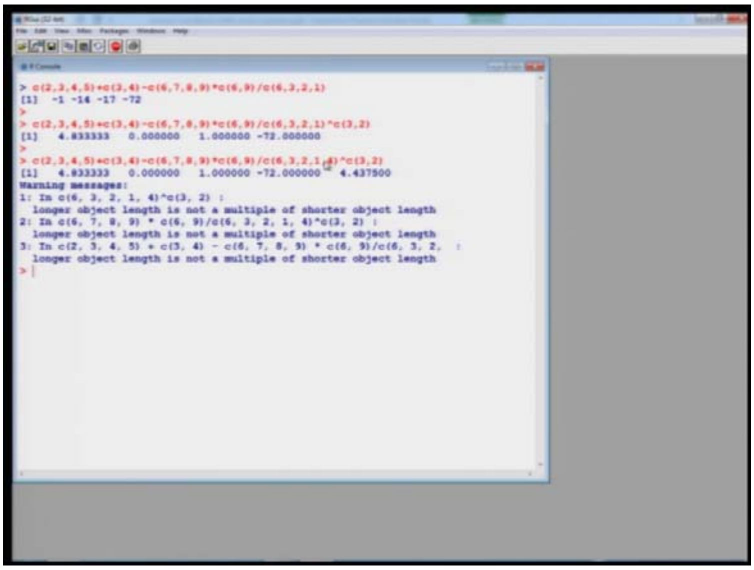
But if I try to take here say one more element such that this is not exactly a exact multiple of the total number of elements in the data vector then there is a warning message, right. So, this is what I meant when I was saying that if the number of elements in the 2 data vectors are not of exact multiple length, then this warning message will come, ok.

Now, you have learnt this thing, suppose if I want to operate the subtraction. So, as we have discussed the operator for the subtraction is the same the classical minus sign, but we have to be careful. So, if I try to take here the a data vector here c 2, 3, 4 and here 5 and if I try to subtract every element by here 1.

So, you can see here the minus sign is being operated on each and every element 2 minus 1, 3 minus 1, 4 minus 1, 5 minus 1 and similarly if I try to take here say here data vector say c 1 comma 2. So, you can see here that what will happen the answer is 1, 1, 3, 3 why? Because 2 and 3 here are subtracted with 1 and 2, then 4 and 5 are subtracted with here another say once again with 1 and 2. So, the answer will be 2 minus 1, 3 minus 2, this is 1 1 and 4 minus 1, this is 3 and 5 minus 2 is here 3.

And similarly, if you try to take some more elements such that they are not exactly multiple. So, you can see here this 2, 3, 4 are being subtracted by 1, 2, 3. So, the answer here is 2 minus 1 which is 1, 3 minus 2 is 1, 4 minus 3 is 1 and after this I have here 5 minus 1 which is here 4, but after that 2 and 3 are not finding any number which has to be subtracted.

(Refer Slide Time: 21:02)



So, that is why it is giving you here this warning message and you see, these things can be, this data vector can also be operated with the BODMAS rule also. For example, if I try to take here say a data vector means addition with say here some number 3, 4 and with subtraction here see here 6, 7, 8, 9 and with say here multiplication with say here 6 and here see here 9 and division with here say here c say 6, 3, 2, 1 and so, on.

So, you can see here this number will be coming. So, the same BODMAS rule whatever you have learnt the same BODMAS rule will be extended to data vectors also. So, there is no issue at all right and in fact, in this BODMAS rule if you want to add here some power operator also, that can also be operated like as a I now I am adding say hat c 3 2.

So, you can see here means again you have the you have a proper result, yes, that you have to see what exactly it is doing because I am just typing here, my objective is to show you that BODMAS rule is still there. And in this case also if you try to violate the rule of exact number of elements, for example, if I try to eliminate say add here one just one number here. Say, you can see here I have added one more number here 4, so that the things are not exactly multiples you can see here that you are getting here some warnings which are saying that something wrong is happening you have to be careful, you have to be watchful and try to see what is, whatever is happening is that really you wanted or this is happening by mistake due to in some programming or in writing the expression.

So, this is what I wanted to convey regarding the mathematical operation of a scalar and as a vector. So, right.

So, now, I will stop here and I have tried my best to give you an overview of the different mathematical operations with respect to a scalar, with respect to a data vector in the R software.

You can see here that this R has a very good capability which is quite different from other software which is the operation with this data vectors and you will see that in statistics, there are many situations where this these data operation will make your life simpler.

For example, if you simply want to know the sum of squares of the number that is pretty simple, you simply have to write down the data just hat 2 and then use a function sum which we am which I am going to discuss in the next lecture. So, just in one line you can write the entire thing.

And then remember one thing this data vector can be stored as a variable also for example, if I say x is equal to c 1, 2, 3, 4, then x square will also be c 1, 2, 3, 4 hat 2. So, all this operation whatever I have shown you here quickly and briefly, they can be extended to a variable also and that is one of the big advantage of using R.

So, that the mathematical computations become easy, but still once again I would say, I have done my job, but your job is pending ,unless and until you practice it you try to settle down this concept inside your brain, you will not be able to understand them you will not be able to use them.

Why? Because only that knowledge is useful which is inside your brain, the knowledge which is written only inside the books, it is helpful for you to understand the subject to gain the knowledge, but unless and until you settle down you store it inside your brain, you will not be able to use it intelligently.

So, you practice it and I will see you in the next lecture some more topics some more details related to the R software. Till then good bye.