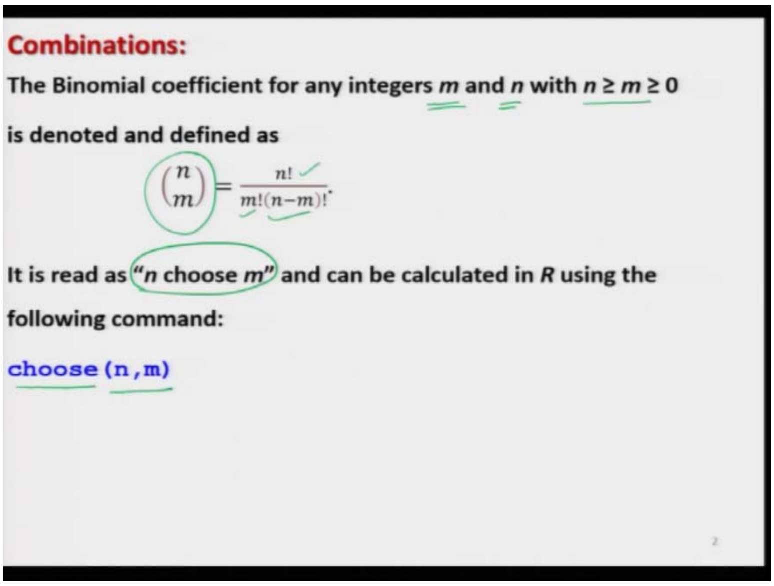
Essentials of Data Science With R Software-1.
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture 16
Basic Principle of Counting-Combinations

Hello friends. Welcome to the course Essentials of Data Science with R Software. And you can recall that in the last lecture, we started learning some basic rules for the counting. Why? Because counting is going to help you in the computation of different types of probabilities. So, we are going to continue on the same topic in this lecture and we are going to now understand combinatorics. If you remember you have done combinations in your class 10 or class 12, use used to write a command say n C r or n choose r.

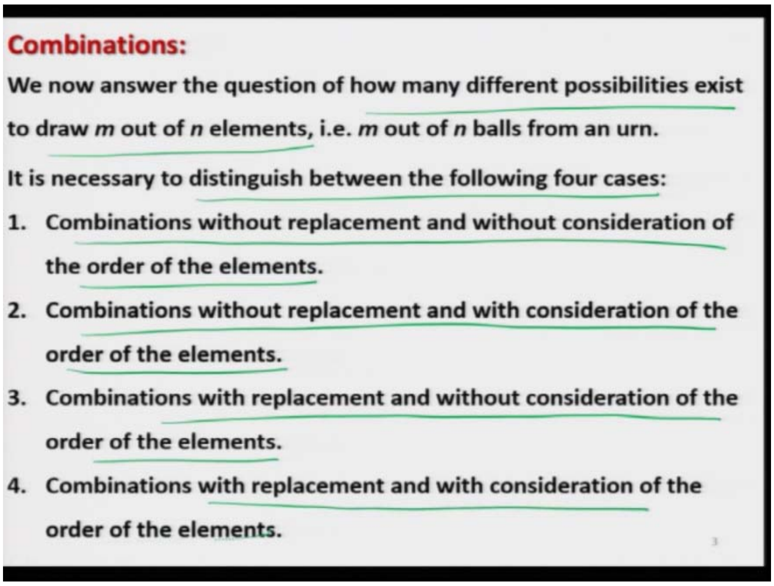
And that used to help us in giving different types of outcome related to the counting process. So, the same command I am going to use here today. And now, how this can help us in counting different types of outcome, that is our objective which we are going to discuss in this lecture. So, let us begin our lecture and we try to understand the combinations.

(Refer Slide Time: 01:25)



So, now you know that the binomial coefficient for any two integers say small m and small n, where they say n is greater than equal to m and both are greater than equal to 0, is indicated and defined as like this symbol, this is called as here n choose m, and the value of this expression is factorial n divided by factorial m into factorial n minus m. So, now, in case if you try to calculate this function in R, then there is a direct command which is choose c h o o s e all in small lowercase alphabets and inside the parenthesis you have to write n comma m.

(Refer Slide Time: 02:12)

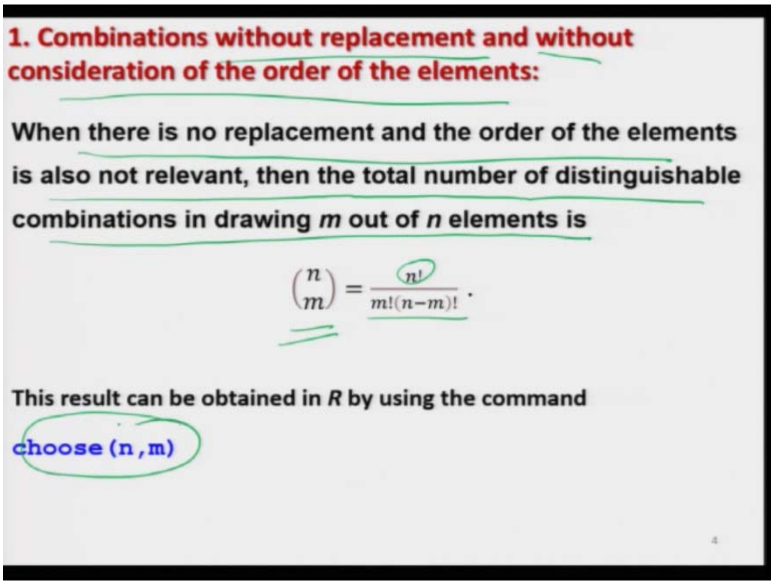


So, now, when we want to understand the application of this n choose m, then we have different type of situations. So, now we try to answer the question of, how many different possibilities exist to draw m out of n element. That we for example, if there is a box, there is an urn in which there are n balls and you want to choose say small m number of balls from that box or from that urn. So, how to get it done?

Now, when you want to do it, there are different possibilities that may occur. So, we are trying to distinguish between the following four possibilities following four cases. The first case we are going to consider here is this combination without replacement and without consideration of the order of the elements. What is the meaning of replacement and order that we already have understood in the last lecture. So, there is no need to explain it again.

The second case will be the combination without replacement and with consideration of the order of the elements. And similarly, the third and fourth options are the combination with replacement and without consideration of the order of the elements and with replacement and with consideration of the order of the elements.

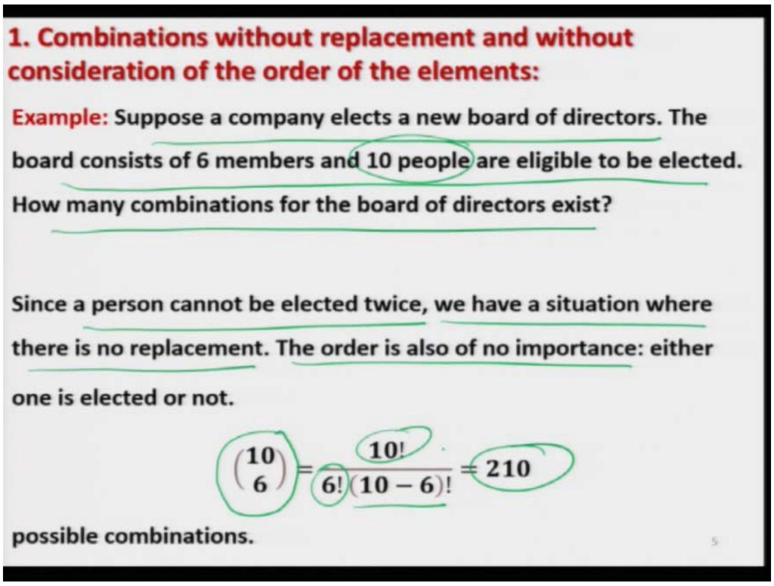
(Refer Slide Time: 03:33)



So, we are now going to learn that under these four different types of cases, how are you going to count the number of possibilities. So, first we try to understand the combinations without replacement and without consideration of the order. So, when there is no replacement, and the order of the element is also not relevant, then the total number of distinguishable combinations in drawing m out of n element is given by n choose m, that we already have done in your class 10 or class 12.

And this value is factorial n divided by factorial m into factorial n minus m and in our software, this can be computed by this command choose n m. Now, let me try to take an example to explain you how these things are going to be applicable.

(Refer Slide Time: 04:12)



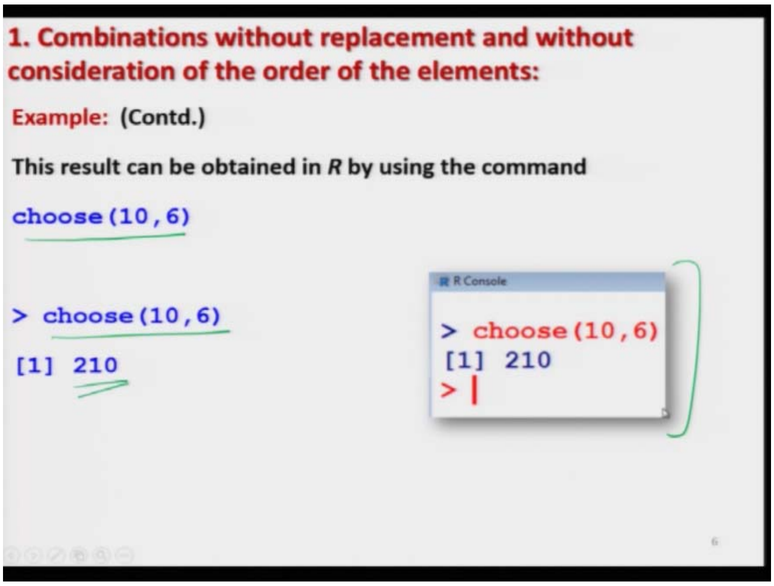
Suppose a company elects new board of directors, the board consists of 6 members and 10 people are eligible to be elected. So that means out of those 10 people, 6 members asked to be elected to be on the board. So, now the question is how many combinations for the board of directors are possible or they exist?

So now, if you want to compute it, I would say simply you have to use some common sense also that how the things can really happen. So, we know that since a person cannot be elected twice, so we have a condition here we have a situation here, where there is no replacement as possible. So now the condition of having no replacement is satisfied.

Now, and the order is also of no importance, whether somebody is elected in the first place or in the second place or in the third place, there has to be only 6 board members. So, well, whether they are chosen in the first row, second row, third row, that does not make any difference. So, right.

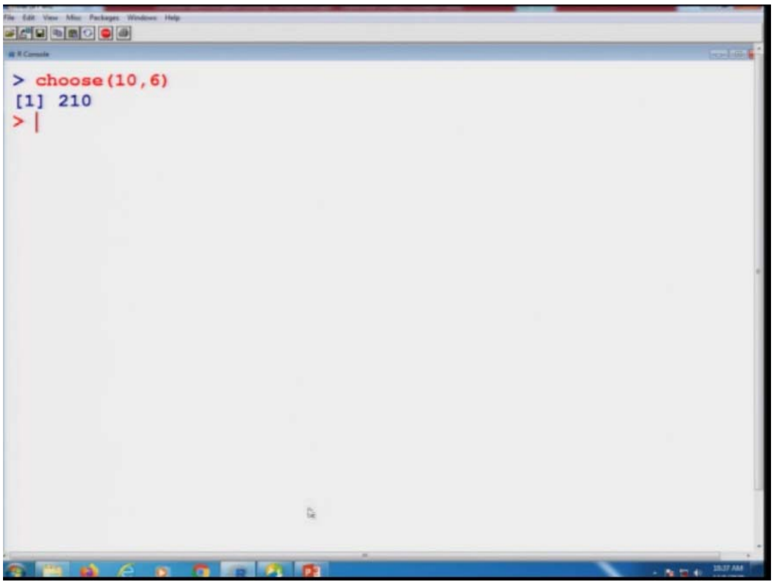
So, in this case, I can choose the number of possible combinations by computing the function 10 choose 6 means 6 out of 10 and then I can use here the formula factorial 10 upon factorial 6 into factorial 10 minus 6 which is 4 and it will come on to be 210. So, there are 210 possible combinations.

(Refer Slide Time: 05:40)



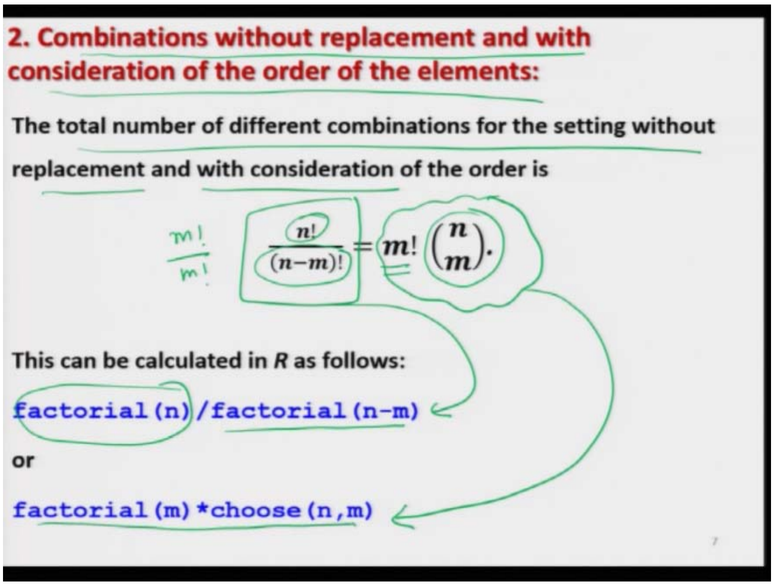
And the same thing you can also do in the R software just by using the command choose 10 6 and if you try to do it here, this value will also come out to be 210. And I will try to show you it on the R console itself that whether this choose 10 6 works or not.

(Refer Slide Time: 05:58)



So, you can see here choose 10 6, and this is coming out to be 10. So, there is no issue, means you can very easily and the computation of this choose function.

(Refer Slide Time: 06:08)

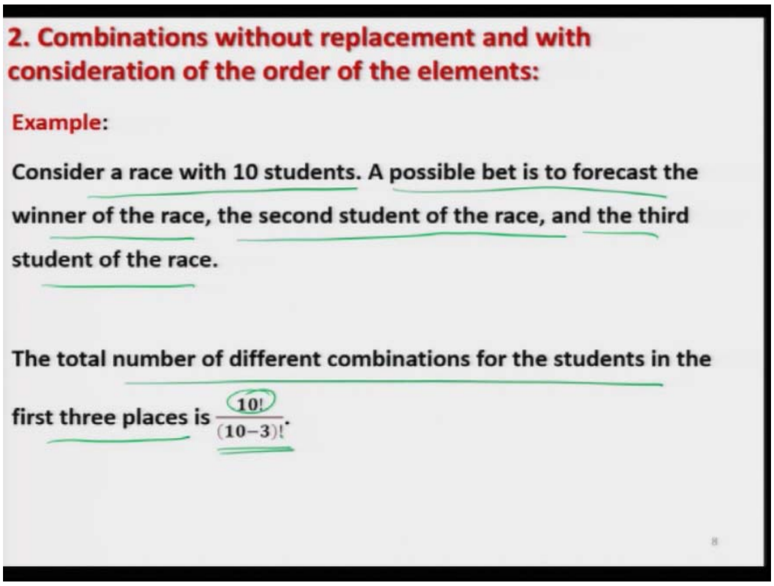


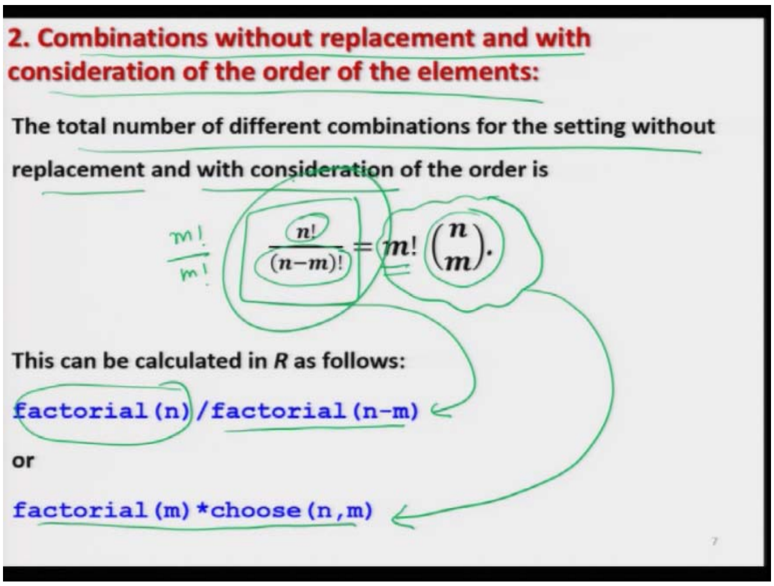
So, now I come to the next topic, where we are trying to find out the combinations which are obtained without replacement and with consideration of the order of the elements. So, in this case, the total number of different combinations for the setting without replacement and with consideration of the order is factorial n upon factorial n minus m. And so, in case if you try to multiply and divide by the quantity factorial m, then it will become here factorial m into n choose m.

So, now, in case if you want to compute this quantity in the R software, you have two options, either you try to consider the quantity this one on the left-hand side of the bracket. So, this can be computed by factorial n divided by factorial n minus m, or in case if you try to compute the quantity on the right-hand side, both are the same thing actually.

So, this can be written in the R software as factorial m into choose n C m. So, this is how you can compute it without any problem in the R software and you know how to compute factorial and how to choose R. So, it will not be a very difficult problem for you.

(Refer Slide Time: 07:19)

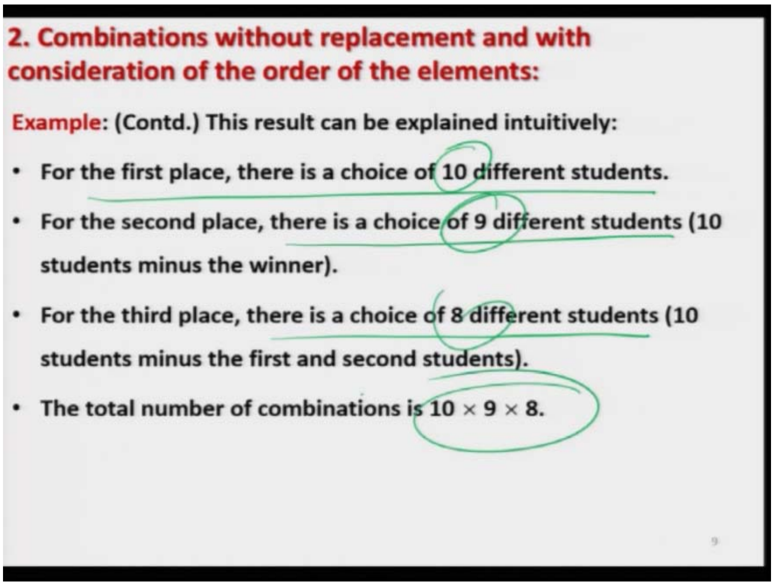




Now, look let me try to give you an example that under what type of condition this is going to be useful. So, we consider a race with 10 students and a possible bet is to forecast the winner of the race, that who is going to be the first, who is going to be the second in the race, and who is going to be there on the third place in the race.

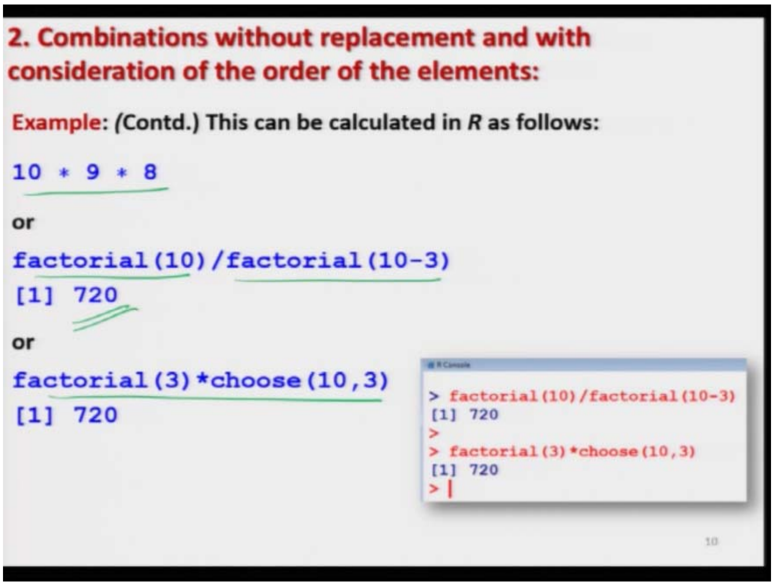
So, in this case, the total number of different combinations for the students in the first 3 places will be factorial 10 divided by factorial 10 minus 3, that is the same formula which you have done here, factorial n upon factorial n minus m.

(Refer Slide Time: 08:00)



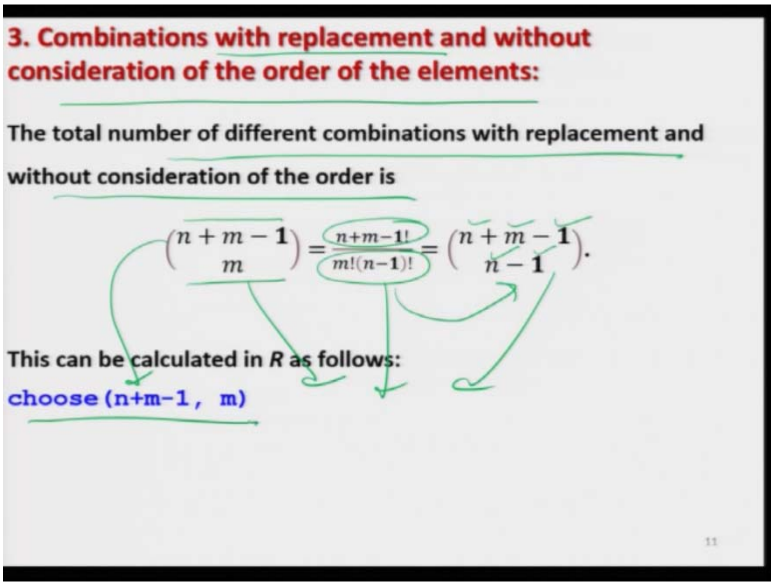
Now, have you tried to understand that how this result is coming out to be. So, I can explain you very quickly, that for the first place, there are 10 students who can occupy the place. So, there is a choice of 10 different students. Now, the first place is occupied. Now, for the second position, there is a choice of 9 different students, because one is to and out of 10 has already been placed in the first position, and then for the third place, two students already have occupied the first and second position. So, there are 80 students left here who can occupy the third position. So, total number of combinations will be this 10 into 9 into 8 which is obtained here like this and, and this is what we have obtained here.

(Refer Slide Time: 08:42)



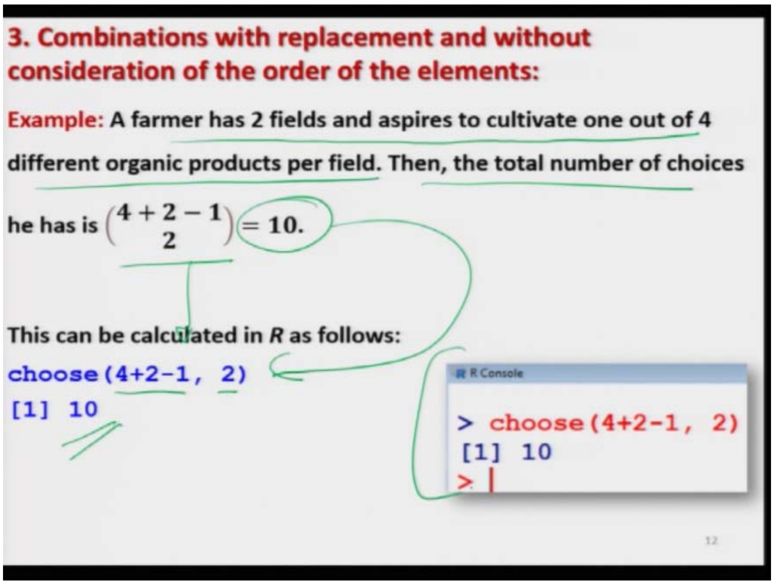
And if you want to calculate this thing that is pretty simple, either you simply multiply 10 into 9 into 8 or you try to write down factorial 10 upon factorial 10 minus 3, it will come out to be 720 or if you try to choose the factorial 3 into choose 10 say 3 that will give you the same value here 720, that will not make any difference. So, you can see here it is not difficult at all.

(Refer Slide Time: 09:03)



Similarly, in that third case, we try to consider the combinations with replacement and without consideration of the order of the elements. So, the total number of different combinations with replacement and without consideration of the order is given by n plus m minus 1 choose m. And if you try to simplify it, this will come out to be something like factorial n plus m minus 1 divided by factorial m into factorial n minus 1 and if you try to simplify this can be written as n plus m minus 1 choose n minus 1. And if you want to compute this quantity, you have three options. Either you can use this form, this form or this form. So, I am just telling you that if you want to choose this form, this can be computed by choose n plus m minus 1 comma m. This command will give you the final answer.

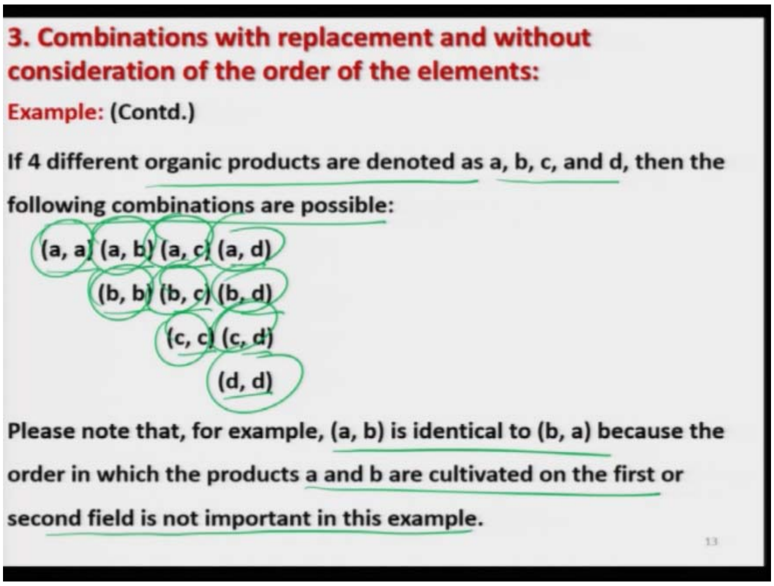
(Refer Slide Time: 09:58)



For example, if I try to take a very simple example, to explain you that under what type of condition this is going to be used, then you can consider a situation where a farmer has 2 fields and aspires to cultivate 1 out of 4 different organic products per field. Then the total number of choices he has will be 4 plus 2 minus 1 C 2 or 4 plus 2 minus 1 choose 2.

And if you try to solve it here, this will come out to be here 10 and if you want to compute this quantity in the R software, you can simply use this function, this will become here choose 4 plus 2 minus 1 comma 2 and this will come out to be here 10. And this is here the screenshot and you can compute it very easily in the R console.

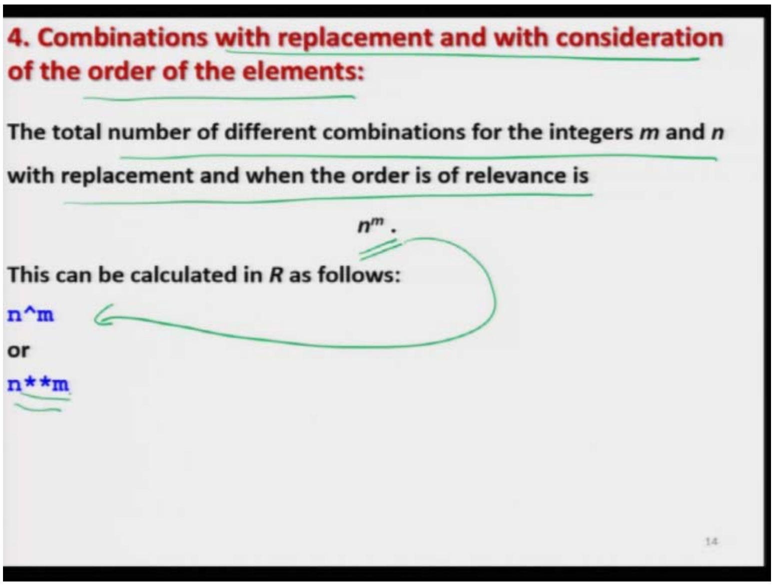
(Refer Slide Time: 10:45)



And if you try to understand that, how this number has been obtained, so I can explain you in a very simple way. Suppose, there are 4 different organic products and and they are indicated by a, b, c, and d. Then then the following combinations are possible, that you try to choose the a,a; a,b; a,c; a,d; b,b; b,c; b,d; c,c; c,d and d,d.

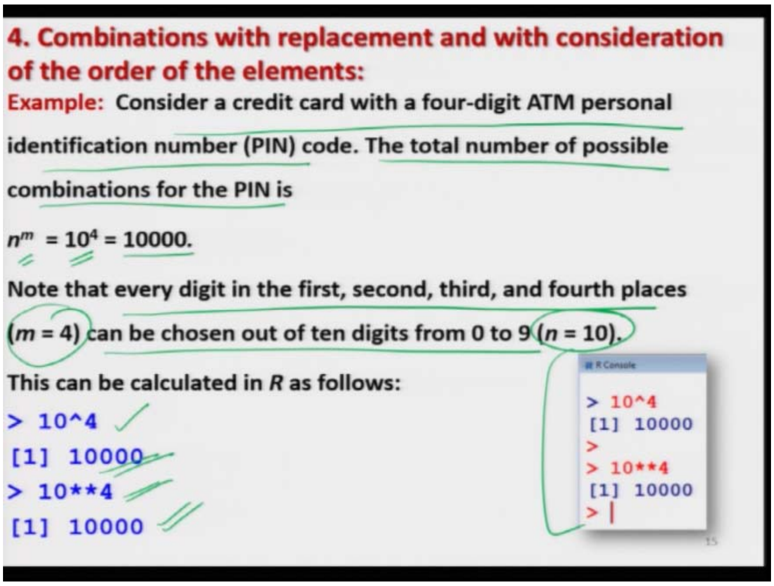
And if you try to count here, how many products are there 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and this is the same thing what we have just computed through our expression. And in this case, what you have to keep in mind that for example, a, b is going to be identical as b, a because the order in which the products a and b are cultivated on the first and second field is not important, it is not given to us. So, now, you can see that obtaining such expressions is extremely simple and they are based on certain logic and counting process, they are not coming from the sky.

(Refer Slide Time: 11:43)



Now, the fourth option. The fourth option is the total number of combinations with replacement and with consideration of the order of the elements. So, that total number of different combinations for the integers m and n with replacement and when the order is of relevant is simply n raised to power of m. And you know this quantity can be computed in R very easily either by writing n hat m or n double star m. That is very simple.

(Refer Slide Time: 12:11)



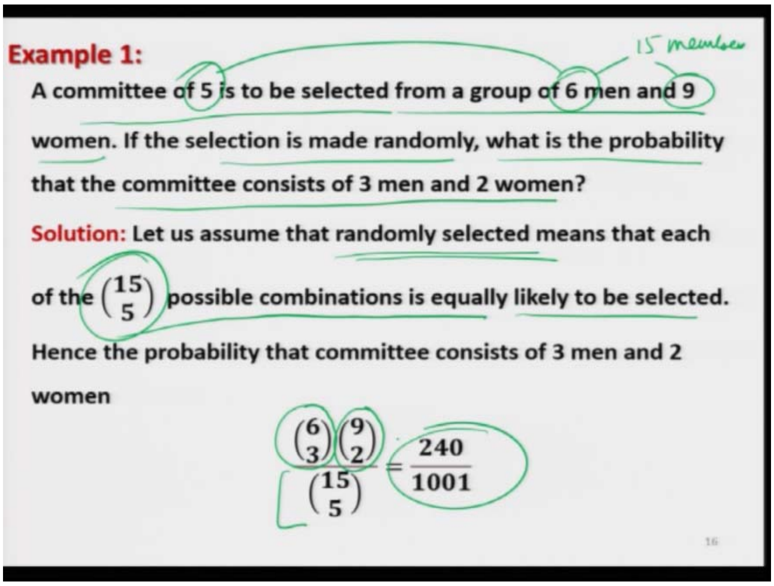
So, let me try to take a very simple example to explain you that under what type of condition this is going to be useful. So, you know that a debit card or a credit card or this card from the bank, they have got a four-digit PIN, personal identification number. And they asked you to choose any four digits to create such a pin.

So, now, we are going to consider a similar example, where we are trying to consider a credit card with a four-digit ATM pin that is personal identification number code. And we want to know, what are the total number of possible combinations for creating such a pin. So, now, this number is going to be very simple- n raised to power of m which is 10 raise to power of here 4 and which is simply here 10,000.

Because what you have to observe here that every digit in the first, second, third and fourth places that is a m is equal to here is four can be chosen out of these 10 digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. So in this case, as well, n is going to be 10.

And if you want to calculate it here, this comes out to be 10 hat 4, which is equal to 10,000 or 10 double star 4 which is the same thing 10,000 and this is the screenshot of the same outcome if you try to operate it on the R console. But now I am sure that these simple operations, you know better than me.

(Refer Slide Time: 13:35)



Now, let me try to take some more example to give you some idea about this counting process. So, I will try to take a couple of examples. Suppose in the first example, I take that there is a committee of 5 members and the 5 members are to be selected from a group of 6 men and 9 woman and the selections are made randomly means there is no choice whether a man or a woman is selected.

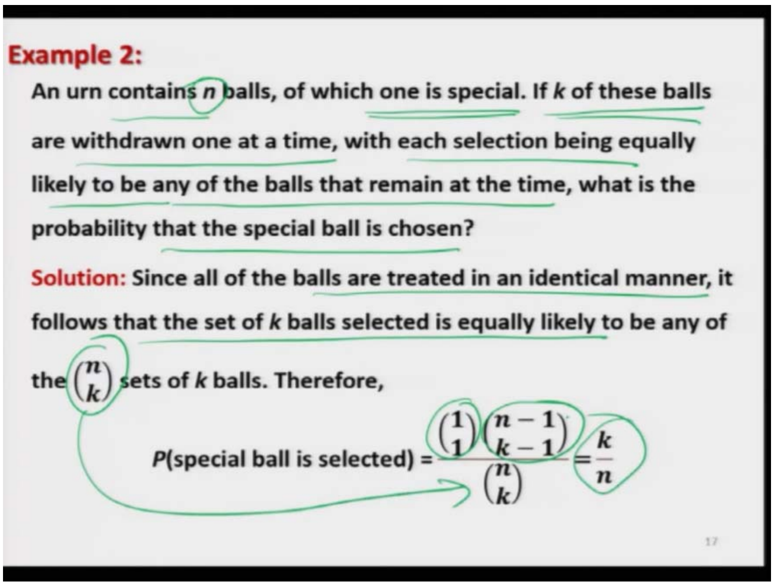
So, now we want to know the probability that the committee consists of 3 men and 2 women. So, now we want to compute this probability. So now, you know that the probability is simply the total number of possible ways divided by the total number of ways. So, the same formula we are going to use here, but now I can show you that how the results what we have learned up to now for the counting they are going to help us in computation of the probability.

So, we assume that as soon as we say that the members are selected randomly, this means that each of the 15 C 5 possible combinations. They are equally likely to be selected. For example, because there are 6 men and 9 women. So, there are altogether 15 members and out of this 15 members you have to choose here, 5 members in the committee. So, this number is going to be 15 choose 5.

So, now, what you want to do, that out of 6 men you want to choose 3 men. So, this number of possible combinations will be obtained from 6 choose 3. Now, there are 9 women and you want to choose 2 women at random. So, the total number of possible combinations will be 9 choose 2 and the total number of ways in which the job can be done that will be the product of the number of ways of selecting men and women.

And the total number of possible cases are 15 choose 5. So, the case is given over here. So, if you simply try to compute it, using the R software also it will give you the value 240 upon 1001. It is not difficult at all you can see. And it is showing you the application of the results what you have just obtained.

(Refer Slide Time: 15:42)

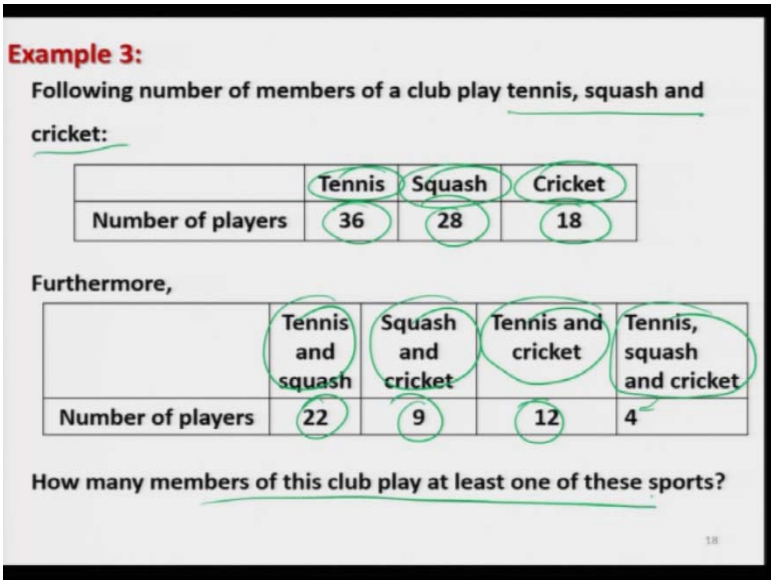


Similarly, if I try to take one more example. So, there is an urn which contains a small n number of balls and out of them, one ball is some special and suppose k number of balls are drawn from this urn at a time and each selection being equally likely that means to be any of the balls that remain at that time.

Now, we want to know what is the probability that the special ball is chosen. So, what we are assuming here that all the balls inside the box they are treated in an identical manner. That means if you try to choose any of the ball from this set of ball, then they are going to be equally likely.

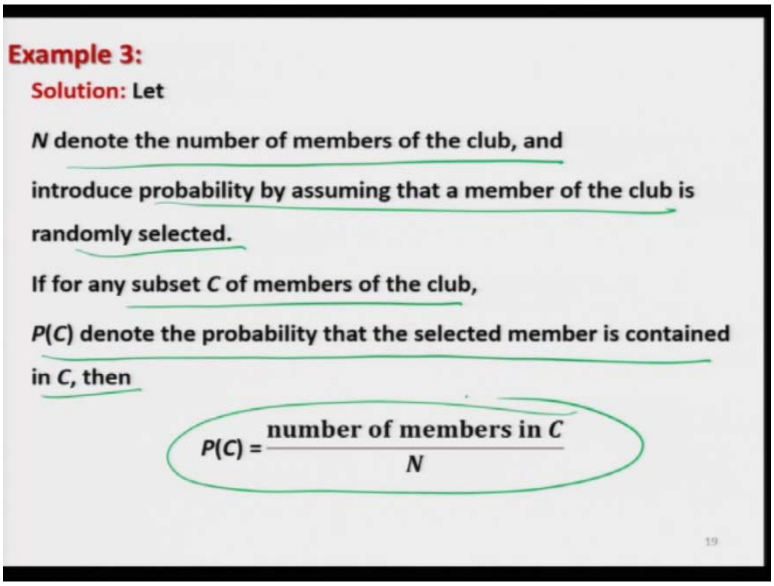
Now, what are the total number of ways in which you can choose k balls out of n balls. That will be n choose k. Now, the probability that the special ball is selected is the total number of ways in which the ball can be selected divided by the total number of ways. So, the total number of ways in which the special ball can be selected is 1 choose 1 into n minus 1 choose k minus 1 and the total number of ways are here and choose k. So, if you try to simplify here, this will come out to be k upon n. So, it is not a very difficult thing for you to solve it.

(Refer Slide Time: 17:09)



And similarly, if I try to take one more example here, that following are the numbers, actually various numbers of members of a club who play the three possible game tennis, squash and cricket. So, the number of players who played tennis game, it is 36; for the squash, it is 28; for cricket, it is 18. And number of players who play both tennis and squash is 22; squash and cricket, this is 9; tennis and cricket, this is 12. And all the 3 games tennis, squash and cricket, this is 4.

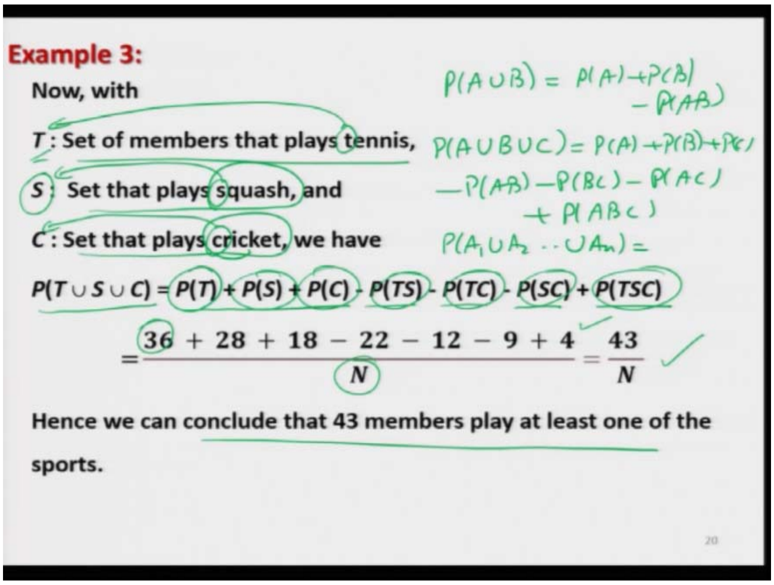
(Refer Slide Time: 17:54)



So, now we want to know that how many members of this club play at least one of these sports. So, now let us try to understand how to solve this problem. Suppose n is indicating the number of members of the club and we try to introduce the probability by assuming that a member of the club is randomly selected. There is no hidden choice or hidden bias or hidden agenda to do to or to select any particular member.

Now, in case if we say that C is the subset of the members of the club, then as per our symbolic notation, this PC is going to indicate the probability that the selected member is contained in the subset C. So, this PC is going to be the number of members in C divided by total number of members in the club that is capital N.

(Refer Slide Time: 18:38)



Now, let us try to consider our numerical values and we try to see all such thing. So now we try to denote by here T the set of members that play tennis, so T is indicating this T and capital S is the set of for those members who play squash. So, this S is indicating here by capital S and C is the set of members, those who play cricket, so this C is indicated by this C. So, now we know that P(T ∪ S ∪ C). This is what we want to find.

Do you remember we had done this result P(A ∪ B), P(A) + P(B) - P(AB). Similarly, in case if I have here 3 events A, B, C, then P(A ∪ B ∪ C) can be written as P(A) + P(B) + P(C) and then minus the combination P(AB) - P(BC) - P(AC) and plus here P(ABC).

And now I have shown you that how you can extend that two events to more than two events. I had told you earlier. But now I have shown you here and actually this result can be extended to any such combination like $P(A_1 \cup A_2 \cup \ldots \cup A_n)$. So first it will come sum then it will be a minus of intersection of two events at a time and then this minus plus minus plus will be getting interchanged and this number of intersections will be increasing.

So, using that formula, I can write down here P(T ∪ S ∪ C as P(T) + P(S) + P(C) - P(TS) - P(TC) - P(SC) + P(T ∩ S ∩ C). And now, if you try to see here this P(T) is the total number of players who plays tennis 36 divided by the total number of members.

So, and then similarly, you can means, you have got all the data the number of players who play squash, cricket or both tennis and squash, tennis and cricket or squash and cricket and those players who plays all the games. So, now you have to simply substitute here these values from this data, what is given you here and you try to compute here this probability will come out to be 43 upon N. Hence, we can conclude that 43 member play at least one of the game.

So, now, we come to an end to this lecture and I would like to stop here. But you see my objective was very simple. I wanted to show you that when you are trying to compute the probability then, you have to count that in how many ways the event can happen. There can be different ways, I have given you only a glimpse, but there is an entire branch of combinatorics which indicates that how this combination can be done.

But my idea was very simple that once you have a logic and you know from the mathematical point of view, how to do the counting, you have to simply write an expression and that expression will be counted by the software that is not your job, because usually these numbers in data sciences are going to be big, but your job is to know how you are going to count, what is the condition under which you can count.

That is just like a doctor and a compounder. The doctor has to decide that based on the symptom which medicine has to be given and the compounder will simply prepare the compound and will give the medicine to the patient. Once you know these things, then computation of probability is not difficult.

The second thing is that once we get a value of the probability how to get an interpretation, now, you know, what is the meaning of the probability value which we are going to get, that is the value when you try to repeat the experiment for a large number of time and then you also have now a axiomatic definition of probability.

So, all those things will make you finally confident that yes, you have got a probability. So, you try to take some examples from the book, try to practice them and I will see you in the next lecture with a new topic in the probability theory. Till then, goodbye.