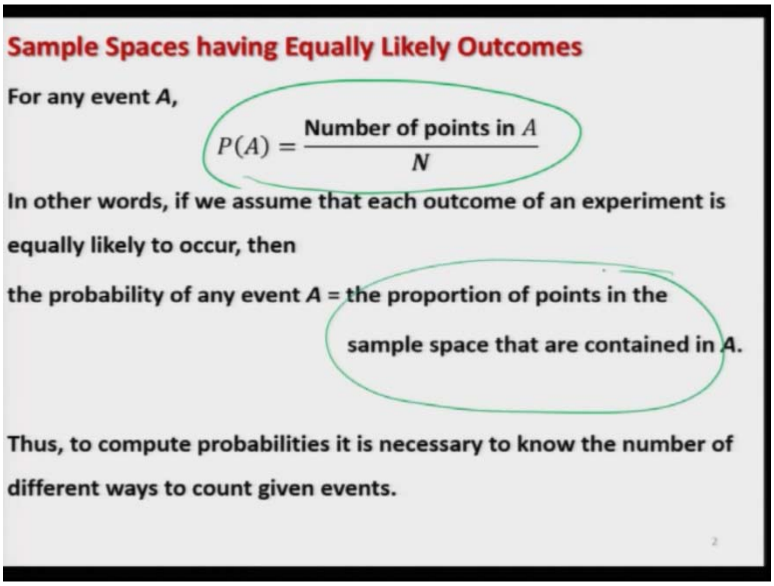
Essentials of Data Science with R Software-1
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture 15
Basic Principle of Counting-Ordered Set, Unordered Set and Permutations

Hello friends. Welcome to the course Essentials of Data Science with R Software 1, where we are trying to understand the topics of probability theory and statistical inference. So, now you can see that we have understood the basic concept, basic fundamentals related to probability theory and finally towards the end of the last lecture, we had discussed that whenever we are trying to compute the probability of an event, the most important ingredient is to count that in how many ways the event can happen, or we simply want to find out the total number of ways in which the event can take place.

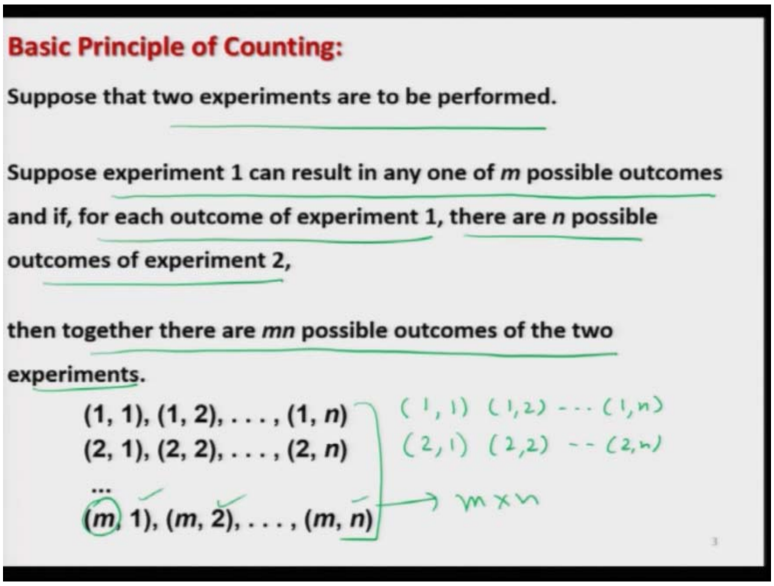
Once you get this thing, you simply have to divide it by the total number of possibilities. So, ultimately the success of computing a probability depends on how successful you are in counting the total number of ways in which the event can favorably occur. So, now, from the mathematics we know that there are various rules for counting, like factorial, combination, etc. So, now in this lecture, we are just going to have a quick review of those concepts. And then we will see how to compute them in the R software, that is our basic idea.

(Refer Slide Time: 02:08)



So, let us try to begin our lecture. So, we try to understand the basic principles like ordered sets, unordered sets and a permutation in this lecture. So, now can you remember, can you recall this slide this was the last slide of the earlier lecture, where we have defined the probability of an event as number of points in a divided by capital letter of N. And we had finally concluded that this probability is simply the proportion of points in the sample space that are contained in A. And now, we need to know that how to count such possible ways.

(Refer Slide Time: 02:39)

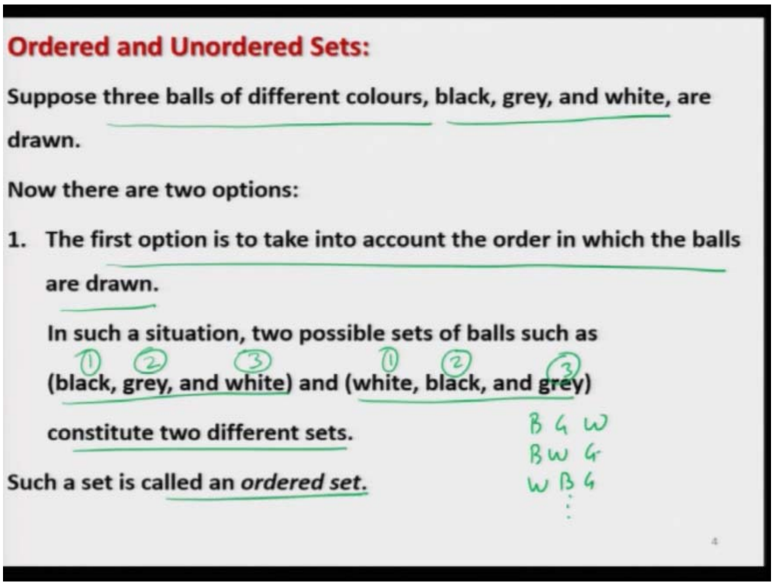


So, now let me try to take couple of such concepts one by one and try to explain you. So, one simple methodology is to just count the total number of ways in which the experiment can really occur. So, suppose, there are two experiments which are to be conducted, which are to be performed and differently these experiments may have different number of possible outcomes. Suppose, that experiment number 1 can result in any one of the possible outcomes that mean there are impossible outcomes.

And suppose, for each outcome of this experiment number one, there are n possible outcomes of experiment number two. So, that way both experiment one and two are interrelated, they are not independent. So, now, we want to know that, how many ways are there in which the events can happen. So, the final outcome is that in such a case together there are m into n possible outcomes for the two experiment, how it will come?

For example, for the outcome number 1 there are say, it can happen with the outcome of the experiment number two say 1 then 2 and up to here 1 comma n. Similarly, for the outcome number 2 of experiment 1 this will be this occurs with the first way of or the first outcome of experiment 2 then 2, 2 then up to here 2, n and if you try to continue here finally, you will have here the mth possible outcome of experiment 1 and that can happen in a small number of ways so 1, 2 up to here n. So, total number of ways if you try to count or the total number of events which can happen, this is here m into n, there are mn ways in which the experiment can be conducted or the outcomes of the experiments can be recorded. So, that is a very simple rule that you have learned earlier.

(Refer Slide Time: 05:03)

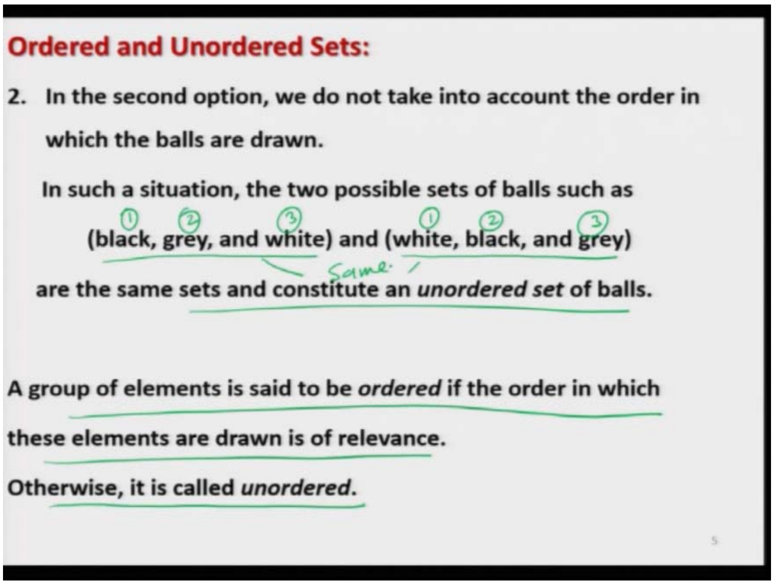


Now, just to give you an idea of the ordered and unordered sets. You see, as the name suggests, ordered, you know, ordered means, you are trying to consider the order in which the outcomes are coming. And when you are trying to use the word unordered that clearly indicates that you are trying to consider an experiment in which the order in which the outcomes are coming is not important.

So, to understand these two simple concepts of ordered and unordered sets, let me try to take a simple example. Suppose, there are three balls of different colors say black, grey and white. And they are inside the box and we try to draw these balls. Now, there are two options. The first option is to take into account the order in which the balls are drawn. For example, you try the first ball, suppose this is black; you try to get the second ball, suppose it comes out to be grey; and then you try to obtain the third ball obviously, that is going to be white because there are only 3 balls.

But this order can be different for example, in case if you try to draw the first ball, this may come out to be white, the second ball that may come out to be black and third ball can be grey. So, you can see here, that these types of different combinations can be made, like as black, grey, white; black, white, grey; white, black, grey and so, on. And these 2 outcomes they are entirely different. They constitute 2 different sets and such a set is called as an ordered set.

(Refer Slide Time: 07:05)

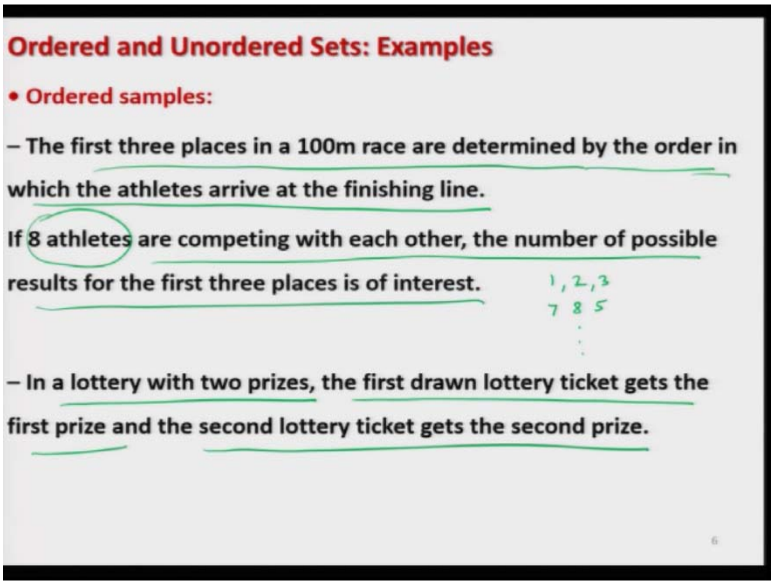


Now, the second option is that we do not take into account the order in which the balls are drawn. For example, you draw the first ball, it comes out to be black; then the second ball, comes out to be grey; and third ball, comes out to be white. Then the first ball, once again it is drawn and the first ball comes out to be white; second, black; and third, grey.

So, you can see here in the first set you are getting black, grey and white and in the second set you are getting white, black and grey. But since you are not interested in the order, so, either you draw the black ball in the first draw, or a white ball in the first draw, it will not make any difference. So, from that point of view, both these sets they are the same and they constitute an unordered set of balls.

So, now I can say a group of elements is said to be ordered, if the order in which these elements are drawn is of relevance, you are trying to consider it. And if not, then obviously, it is unordered. That is as simple as, that is the most simple definition to understand this concept.

(Refer Slide Time: 08:19)



Now, let me try to take some examples of this ordered and unordered samples. For example, you know that whenever there is a race, a couple of people run and when and then finally at the end, we want to know who come first, who comes second and who came comes third.

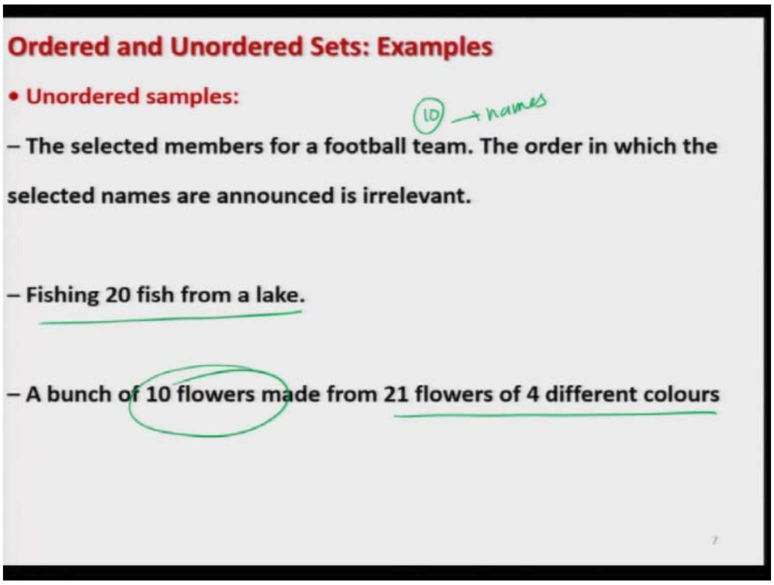
So, now if there are more than one athletes, means obviously if there is only one athlete that will always come first. So, suppose there are eight athletes, and they are trying to take a part in a competition, where they are trying to take 100-meter race. And we want to know the first three places in this race are occupied by which of the athletes. So, the first three places in this 100-meter race that is going to determine that who are winner, but that winner is determined by the order in which the athlete arrived at the finishing line.

Now, you can say here, suppose eight number of athletes are competing with each other they are participating in the race. Then the total number of possible results for the first three places will be of our interest. For example, it is possible that first, second and third number of athletes they come in the first three position or it is possible that 7 number, 8 number or say 5 number athlete they come in the first three positions and so on there can be different ways.

And but in this case, you can see the order is very important, means any athlete who is reaching first that athlete will be given the first price. The athlete who is reaching in the second place, that athlete will be awarded the second price. And the athlete which is reaching in the third place, third position, that athlete will be given the third price. And any of those 8 athletes can occupy any of these three positions depending on that how much time they take in reaching to the goal.

Similarly, suppose our lottery has two prizes, first prize and second price. Now, people purchase these tickets, and then we have to draw two tickets out of those tickets. So whatsoever first drawn, lottery ticket comes out, that ticket will get the first prize. And whatsoever second lottery prize comes at the second position, that will get the second price. As simple as that. So, in this case, you can say that ordering of the events in which they are performed, that is important.

(Refer Slide Time: 11:26)



Now, I try to take an example of unordered sample. Have you ever seen how our football team is framed? Suppose in your college also, there are 200 students or 500 students and suppose out of them suppose there are some student who play the football. So now, the ultimate goal is that suppose that a team has 10 members in the football team. Well, I do not know exactly what is the real number of members in the football team, but I am assuming suppose there are 10 people.

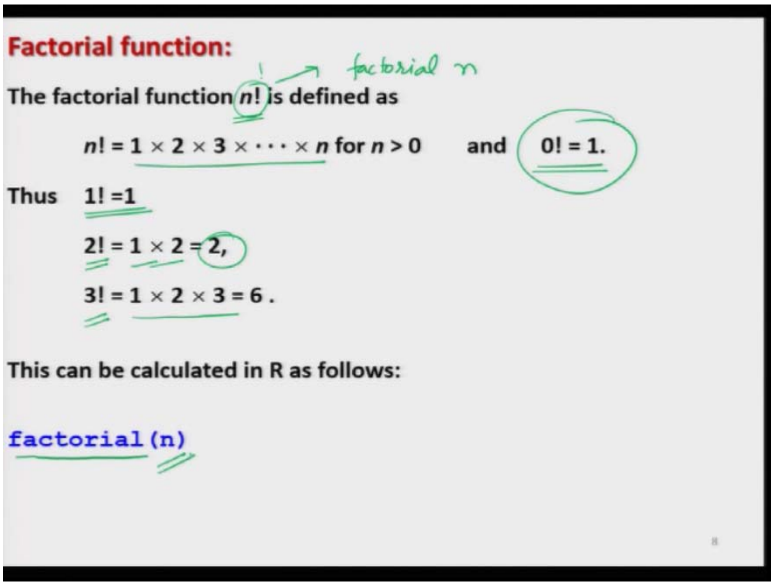
So, now, their names will be collected by a committee and then their names are going to be announced. And finally, suppose 10 names are to be announced. Now, in this case, do you really think does it make any difference that the student whose name is announced at which of the order is really going to make any difference, certainly not.

Whether the name of first student is announced first or the name of the first student is announced in the last that will not make any difference. The only thing here is that the name is announced or not, because names of only those students are going to be announced who are going to be in the team.

And similarly, if you want to suppose a fish 20 fishes from a lake, the person will simply try to put a fishing rod and what say what 20 fishes are caught, they will be a sample. So, in this case, the order in which the fishes are being caught that is not important, that is immaterial, because our objective is to just get 20 fishes.

Similarly, suppose there are 21 flowers of four different colors and we want to make a bunch of 10 flowers, then means I can choose any flower from this from this group of flowers and can create a bunch. So, in this case, whether the white flower is drawn first, or the blue flower is drawn first that will not make any difference, unless and until we have a special request or interest. But in general, in all the three examples, we are not concerned about the order in which the observations are drawn, and that is why these are the unordered samples.

(Refer Slide Time: 14:11)

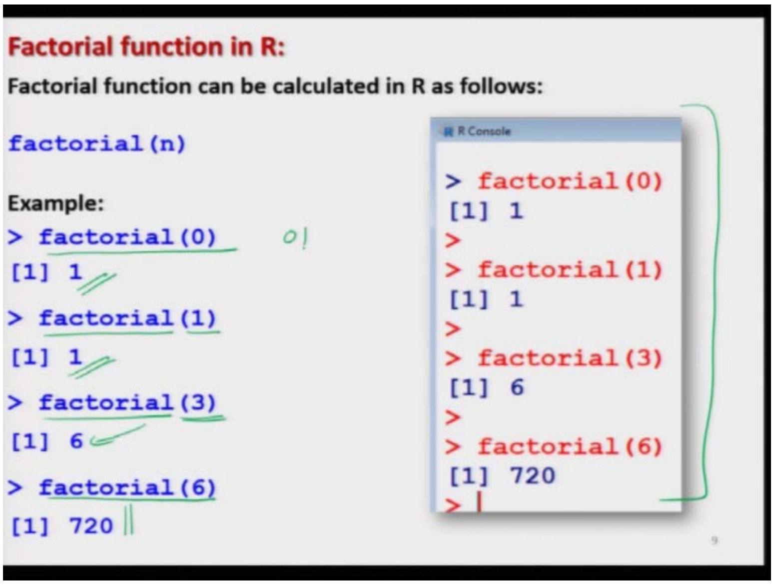


Now, I try to introduce here major that you already know, that is the factorial function. The factorial function, this is defined or indicated as like n and this exclamatory sign. And what is the meaning of this n and this exclamatory sign, this is called as factorial n, factorial n. So, the meaning of factorial n is that all the integers starting from 1 to n they are multiplied together, like as 1 into 2 into 3 up to here n. And yeah, there will be some confusion. So, it has been assumed that factorial 0 has the value 1.

And similarly, if you try to take here the factorial 1 that will also have the value 1. So, that is our assumption that factorial 0 will always take the value 1. So, now in case if you want to find out the factorial 2, the value of the factorial 2 that is simply going to be 1 into 2 which is a 2 and similarly, if you try to find out the factorial 3 that can be obtained by 1 into 2 into 3 which is here 6.

Now, as this number increases the value of factorial and will also become larger. So, in R this factorial function can be computed using the function factorial f a c t o r i a l, and inside the parentheses you have to write the number for which you want to compute the factorial. So, I am writing here, so in general, say factorial n, so n is going to be inside the parenthesis.

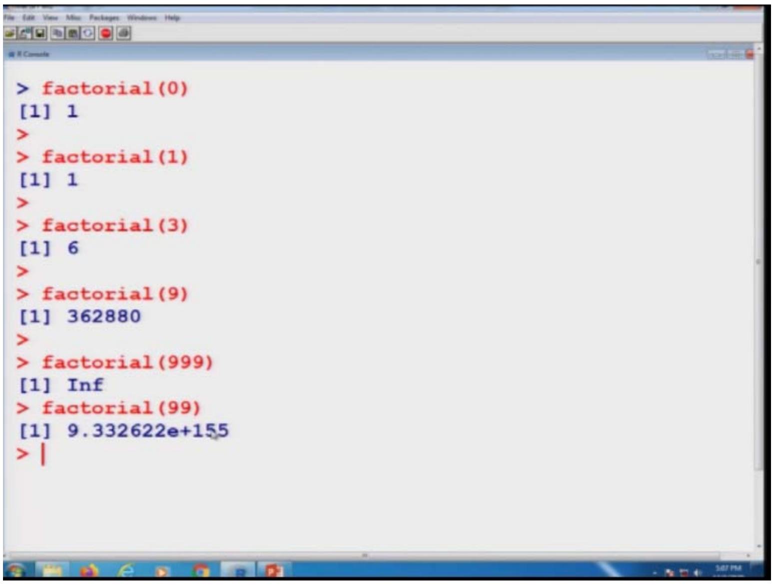
(Refer Slide Time: 16:01)



And now I try to show you some examples and I will try to show you it on the R console also. Suppose I want to find out the factorial 0, the value of factorial 0, like this, so it will be factorial and inside the parentheses this is 0, this will become here 1. And similarly, for the factorial 1, I have to write down here factorial and within parenthesis write 1, this will give you here a value 1.

And similarly, for the factorial 3 also we write factorial 3, the 3 is inside the parenthesis, and it comes out to be 6. And similarly, if you want to know the value of factorial 6, that will be 1 into 2 into 3 into 4 into 5 into 6, which will come out to be 720. And you can see here, this is the screenshot of the same event.

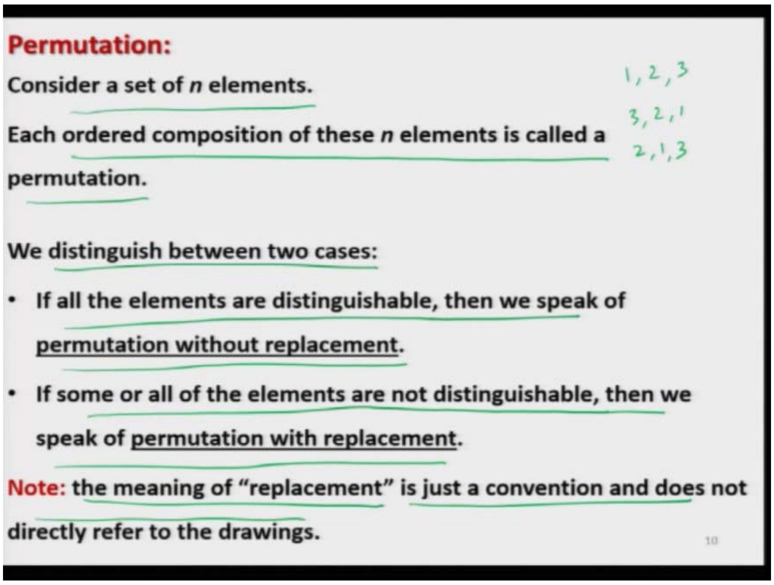
(Refer Slide Time: 16:58)



So now, let me try to show you these things on the R console itself. So now let us come back to R console and if you try to see here, I tried to find out here the value of factorial 0. This is here 1. And similarly, if you want to find out the value of factorial 1, this will again come out to be 1 as I shown you. And similarly, if you want to find out here the value of factorial 3, you simply have to write down factorial inside parenthesis 3, and it will come out to be a 6.

And similarly, if you want to find out here factorial 9, this will come out of here like this. And similarly, if you try to say, find out here, factorial 999, let us see what happens, it will be infinity because the value is very, very large. But if you try to write down here, we will know the value of factorial 99. This will give you 9.33 into 10 raise to the power of 155. You see if you try to calculate these values manually, it will be very difficult but R can do it very easily.

(Refer Slide Time: 17:55)

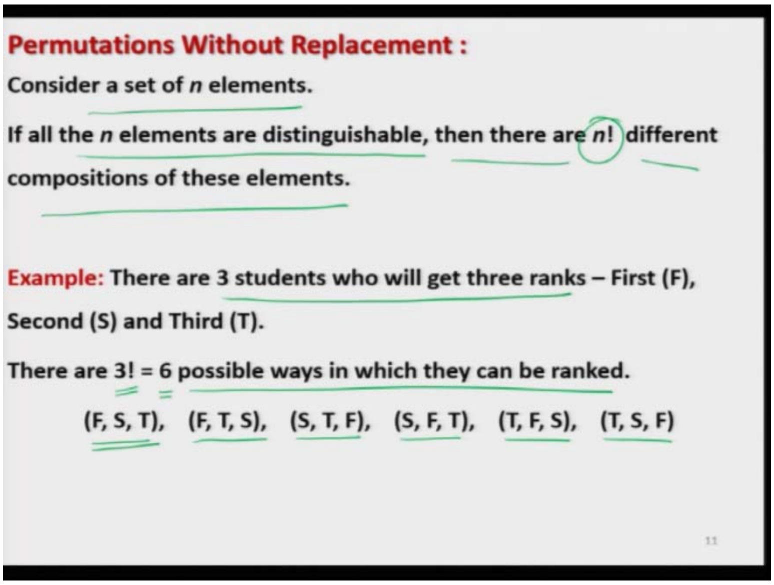


So, now let us come back to our slides and let me try to give you an idea of another tool, what we call as permutation. And I am sure that you have done these concepts in your class twelve or so. So, what we try to do here that we try to consider here are a set of a small n number of elements and each ordered composition of these n elements is called a permutation.

For example, if I say here, I have three numbers here 1, 2, 3, so 3, 2, 1 or say 2, 1, 3. These are simply the permutations of 1, 2, and 3. Now, in this case, we distinguish between the two cases. If all the elements are distinguishable, then we speak of permutation without replacement. And if some or all of the elements are not distinguishable, then we speak of permutation with replacement.

One has to note that the meaning of this replacement is just a convention and does not directly refer to the drawing, means you have to be careful that I am not using this word replacement in terms of simple random sampling with replacement and without replacement. Be careful.

(Refer Slide Time: 19:17)

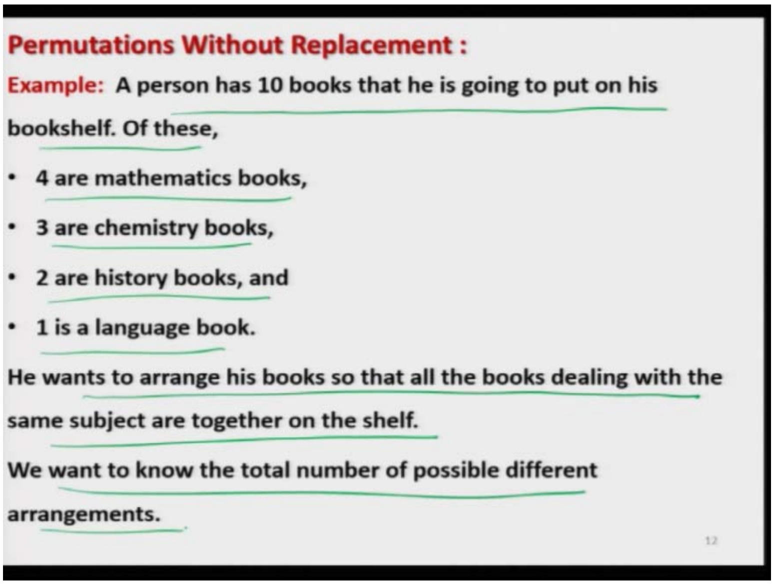


Now, let us try to take an example. But before that, what is the meaning of this sentence that n elements are distinguishable? So, consider a set of small n elements. If all the n elements are distinguishable, then there are going to be factorial n different compositions of these elements. For example, if there are 3 students, who will get 3 rank first, second and third.

Now, what are the different possible ways in which they can get these ranks? It is possible that is to one is getting first rank, two is getting second rank, third is getting the rank number three. Or it is possible that is number 1, 2, 3 gets ranked 3 2 and 1 and so on. So, in order to count such numbers, we can use the concept of factorial and we can write down here that there are factorial 3, which is equal to 6 possible ways or there are 6 possible permutations of these 3 students. What are these, say, first, second, third; first, third, second; second, third, first; second, first, third; third, first, second; third, second first.

And similarly, you can count different ways and now, you can see that whenever you are trying to compute a real probability whenever the data size is very large, there can be last number of such possible combination then such rules will help you in finding out the total number of ways.

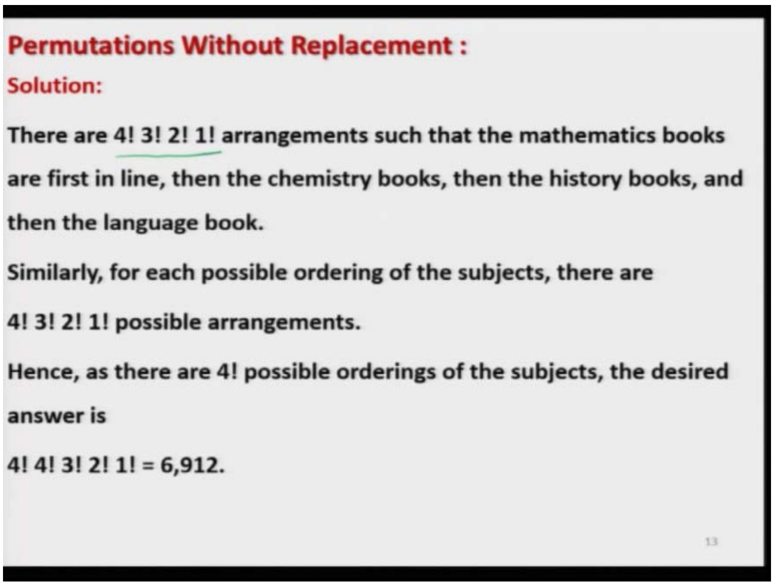
(Refer Slide Time: 20:53)



Let me try to take here one more example. Suppose a person has 10 books and he is going to put those books on his bookshelf. And out of these 10 numbers of books, 4 are Mathematics book, 3 are Chemistry book, 2 are History books and 1 is language book. And he wants to arrange his books so, that all the books dealing with the same subject are together on the shelf.

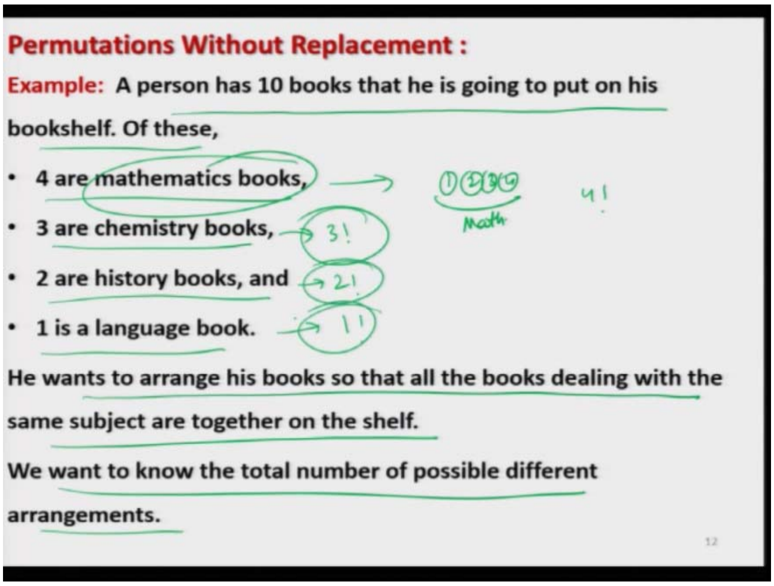
And we want to know the total number of possible different ways in which this job can be done. For example, there can be 4 Mathematics book first, then 3 Chemistry books; or there can be 3 Chemistry book first, and then 4 Mathematics books after that and so on.

(Refer Slide Time: 21:38)



So, in this case, there are factorial 4 into factorial 3 into factorial 2 into factorial 1 possible arrangements as the Mathematics books are in the first line, then the Chemistry book, then the History books and then the language book.

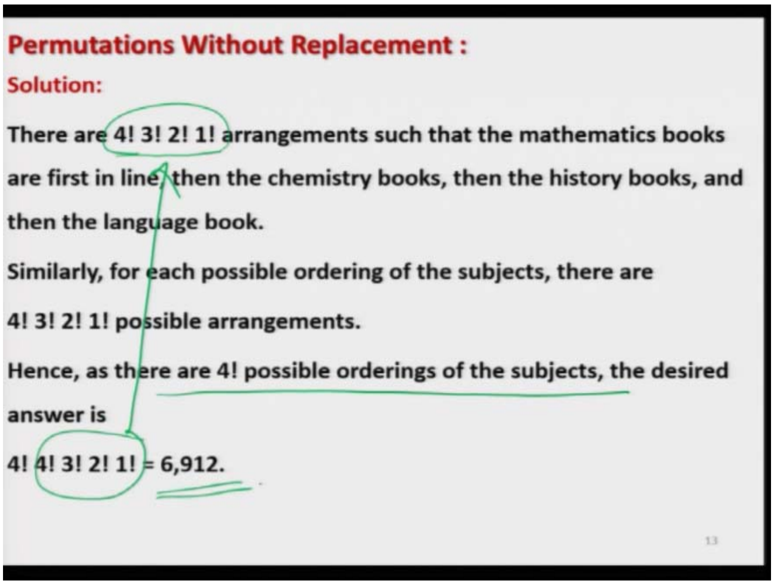
(Refer Slide Time: 21:56)



Because, you can see here, there are 4 Mathematics books. Well, I am not discriminating between the book number 1, 2, 3 and 4 of Mathematics. No. They are not distinguishable, they are same for us. So, that is why there are factorial 4 ways in which the Mathematic books can be arranged. And similarly, for the Chemistry also, we are not discriminating between the book number 1 2 and 3. But we are simply saying there are factorial 3 ways in which the Chemistry books can be arranged. Similarly, there are factorial 2 ways in which the History books can be arranged and there is only one way in which the language book can be arranged.

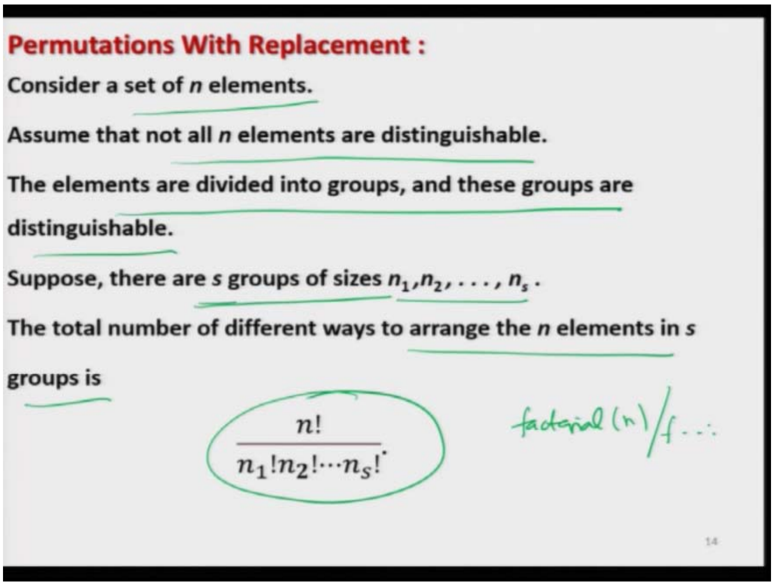
So, the total, if a job A can be done in n ways, the job two can be done in n ways, so the total number of ways in which the job can be done it is, I mean, 2n. That is what we already have discussed. So, using that rule I can say that here that the factorial 4 is the number of ways in which the Mathematics books can be arranged, Chemistry book can be arranged in factorial 3, History book can be arranged in factorial 2, language book can be arranged in factorial 1 ways.

(Refer Slide Time: 23:08)



So, there are total factorial 4 into factorial 3 into factorial 2 into factorial 1 possible arrangement says that the Mathematics books comes in the first line, then Chemistry, followed by History and then finally language. Now, similarly, there are 4 subjects, so, these four subjects can also be interchanged. So, there are factorial 4 possible orderings of the subjects. So, now, the total number of ways are going to be factorial 4 into this number, and that will come out to be 6912 if you want to solve. So, you can see here this is how we can compute different types of possible combinations that we want.

(Refer Slide Time: 23:51)

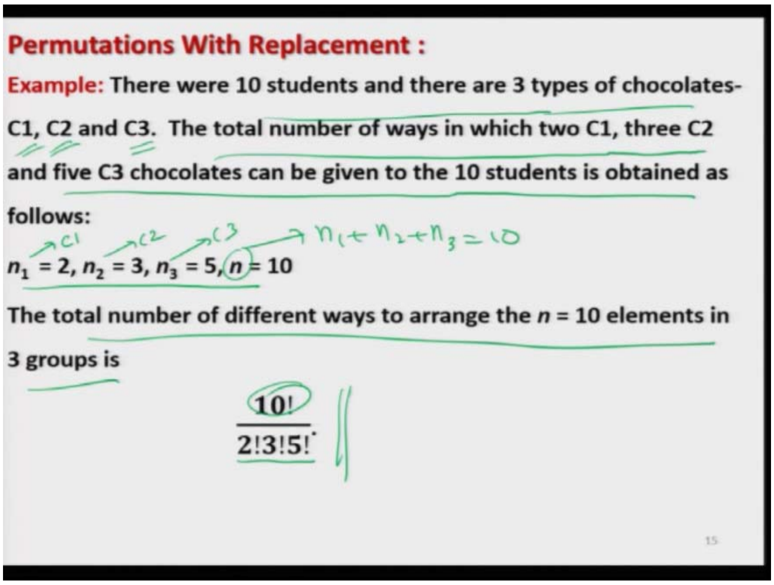


Now, we try to consider the permutation with replacements. So, we try to consider here a set of n elements and we assume that not all n elements are distinguishable and the elements are divided into groups and these groups are distinguishable. For example, if there are 100 students in the class and they are divided into suppose five groups consisting of 20 students each now these groups are distinguishable.

So, in general I can say here suppose, there are a small s number of groups of sizes n_1, n_2, …, n_s. So, the total number of ways, total number of different ways to arrange the small n element in s group is factorial n upon factorial n_1 into factorial n_2 up to factorial n_s. So, this is how you can find it out.

Now, if you want to compute it the R software, it is very simple. You know how to compute the factorial. The command here is factorial say n and you have to divide it by factorial n_1, factorial n_2 up to n factorial n_s, whatever are the numbers and then you can very easily compute it on the R software. So, you can see it is not difficult to compute these values in the R software.

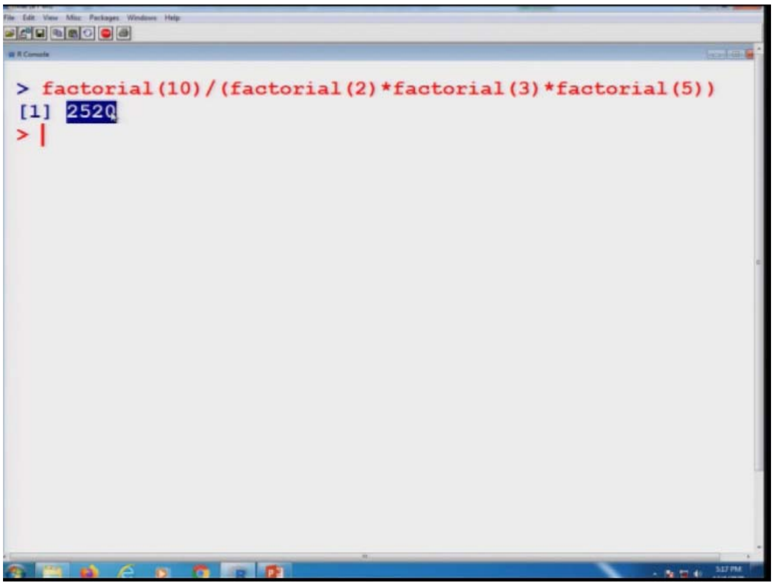
(Refer Slide Time: 25:18)



So, let me try to take an example to explain you this thing. Suppose there are 10 students and 3 are three types of chocolates say chocolate number 1 indicated a C1, then chocolate 2 C2, chocolate 3 C3. The total number of ways in which 2 C1, 3 C2 and 5 C3 chocolates can be given to the 10 students is obtained as follows.

Now, in this case, you can see here that, we have to first indicate the these numbers. So, we try to assume here that n_1 is equal to 2, n_2 is equal to 3 and n_3 is equal to 5 and n is obtained by n_1 plus n_2 plus n_3 which is equal to here 10. Because this n_1 is going to indicate the total number of chocolates of type one C1, n_2 is for C2 and n_3 is for C3. So now, I try to simply use this formula and try to obtain the total number of different ways to arrange say 10 elements in three groups, which is factorial 10 upon factorial 2 into factorial 3 into factorial 5.

(Refer Slide Time: 26:38)



Now, if you asked me to compute it on the R software, yes, I can very easily compute it. So, let me try to show you here how I can do it in the R software. I will simply write down here factorial 10 divided by factorial 2 into factorial 3 into factorial 5, that is all. And there are 2520 number of ways.

Now, if I asked you that, can you really count them manually. It is very difficult. But using the software you can see that these things can be obtained very easily. So, now, let me come to an end to this lecture, but, I have not completed these topics. In the next lecture, I will come up with some more topics that how you can count, but in case if you try to see in this lecture, we have not done anything new.

These are the things that you already have done earlier. But my job was to inform you that yes, these are the things which are going to be useful for you, when you are trying to compute different types of probabilities. And secondly, in how are you going to compute these things in the software. That was my basic objective of giving this lecture to you. So, you try to have a look, try to practice it, and I will see you in the next next lecture with some more rules of counting. Till then, goodbye.