Professor Swagato Chatterjee
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture 15 – Segmentation, Targeting and Positioning (Contd.)

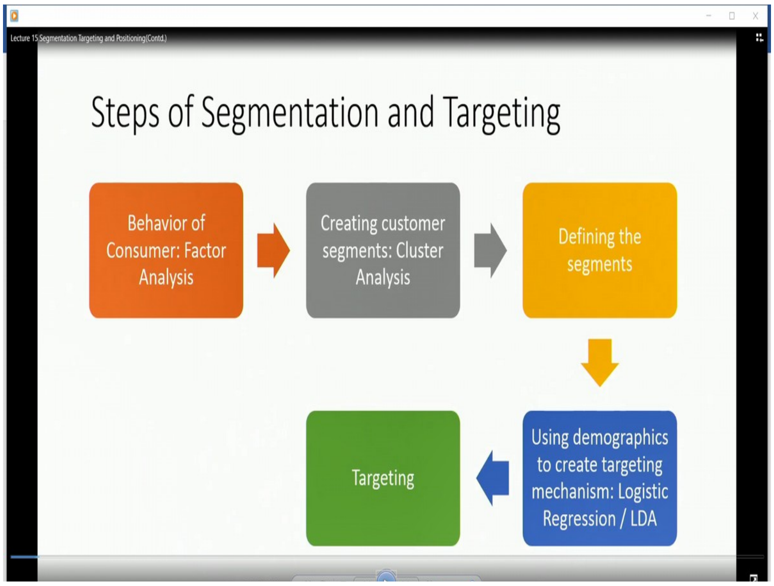
Hello everybody! Welcome to Marketing Analytics course. This is Dr. Swagato Chatterjee from Vinod Gupta School of Management, IIT Kharagpur who will be taking this course for you. We are in module three now and we are discussing segmentation, targeting and positioning.

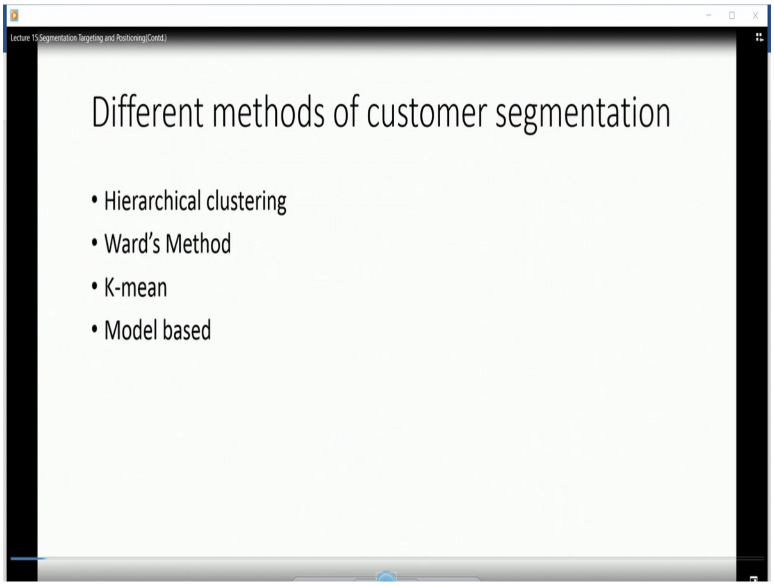
So in the last video, we have discussed at length about what is segmentation and how to do it, what is targeting and what is the utility of targeting and how to position your product when you target a market.

(Refer Slide Time 00:45)

So here, in this presentation, we will go ahead from whatever we have done in the last class.

(Refer Slide Time 00:52)





So first thing that we will discuss about the steps of segmentation, targeting and positioning and these are the five steps that are there. Behavior of consumer and we do it through factor analysis – that is the first step. And then we do some…. And factor analysis will have this kind of methods. Sorry, this is the second step. This is cluster analysis step.

So first we do factor analysis and then... Why we do factor analysis? Because there are lot of behavior that are involved and sometimes those behavior has to be clubbed together so that some similar kind of behavior comes up. And the second step is cluster analysis where based on the behavior of the customer we try to find out which customers have similar kind of behavior and which customers do not have similar kind of behavior.

And some of the classic methods that we apply to see that which kind of customers are coming close to each other and which type of customers are going away from each other, we use, these are some of the methods like: Hierarchical clustering. Ward's is one of the part of hierarchical clustering method and there is another algorithm called K-mean algorithm and then model based algorithms as well.

So second step is, so first step is to combine the behaviors to find out certain behavior which is, for example let us say price sensitivity. Price sensitivity is a behavior or probably sometimes the attitude of the customer towards the price. Now that can be seen as the behavior. You cannot actually ask all the customers what is their price sensitivity but you can capture that ok, this is the customer who only buys during the sale season or he buys only when there is a particular, some amount of offs are going on or he always uses coupons. He cuts down coupon from newspapers or probably certain kind of coupon codes from different sources and he spends time on that, finding out the sources of coupons and etcetera and then he makes the purchase.

Or he is not very happy to give let us say delivery charges in an ecommerce setup. He actually looks for options how to get away with the delivery charges. Now all of this behavior sometimes are related to a similar pattern of that particular consumer or similar attitude of that particular consumer which is expressed in their behavior which is let us say price sensitivity.

In the other case, there can be let say brand awareness. So a guy who only purchases major brands and who search for those brands or who generally, even when the sales promotion of the non-branded product is going on or some other brand is going on, he will be loyal to that particular brand. So all of these things might come together as some single behavior as well. So there can be so many behaviors, consumer makes so many behaviors in a retail store or in ecommerce setup or different places. So we have to do first factor analysis and we will discuss about that in a different part of the course – how to combine them and create certain meaningful characteristics of the customer.

Now after we found out different meaningful characteristics of the customer, we are trying to club those customers based on whether the meaningful characteristics are matching with each other or not. And if they match with each other, if customer A and customer B have similar meaningful characteristics then they will be close to each other and they will be kept in the same bucket and if this one set of customers, one bucket of customers is very away from the other bucket of customers then that is a good segmentation. That is something that we try to achieve.

And the methods are hierarchical clustering, Ward's method, K-mean and say other methods also and a combination of them. So once you have created, still now you remember that we have not talked about demographics. We have created all these clusters based on similar characteristics which is behavioral characteristics. So we are not talking about demographics. But, you have to, once you have created two segments, three segments, you have to name them, you have to name what that segment is.

So let us say, I found out there are certain people who are let us say price aware or price conscious and then there are certain people who are like some other kind of behavior comes up. Now who are they? What is this person? If a new customer comes in my retail store, how will I identify that whether this guy is segment A or segment B or segment C? So for that, we have to name the segments.

So for that, we use something called defining the segments, that is the third step. Defining the segments which is the third step and then rather than defining, we actually try to find out how I can predict the segments. And generally, we use methods such as let us say, we use methods such as LDA – Linear Discriminant Analysis and there is a method called Multinomial Regression.

So at this moment, if you are with me, we will be discussing various methods, but you can go pause the video and probably search, I will share certain links also, you will be finding it in the description or additional materials where certain links will be given where you can read about what factor analysis in general is, what linear discriminant analysis in general is or what multinomial regression is. These are the kinds of topics which will be covered in a which should be covered in a business analytics kind of course. So I am not covering the nitty-gritty of the topics, what these things are, what is the algorithm, what are the findings and etcetera, but the part that we will be needing in our analysis, we will cover that.

So, now given that, what we will do, LDA and multinomial regression is nothing but trying to predict the segment of a customer based on certain demographics. So, this is the first part where demographics come in when you are trying to target, when you are trying to create a profile, create a name, create an identity of a particular segment. You are creating a human being's, how the human being will look like, that is when you are trying to create a segment and that is where the demographics comes in.

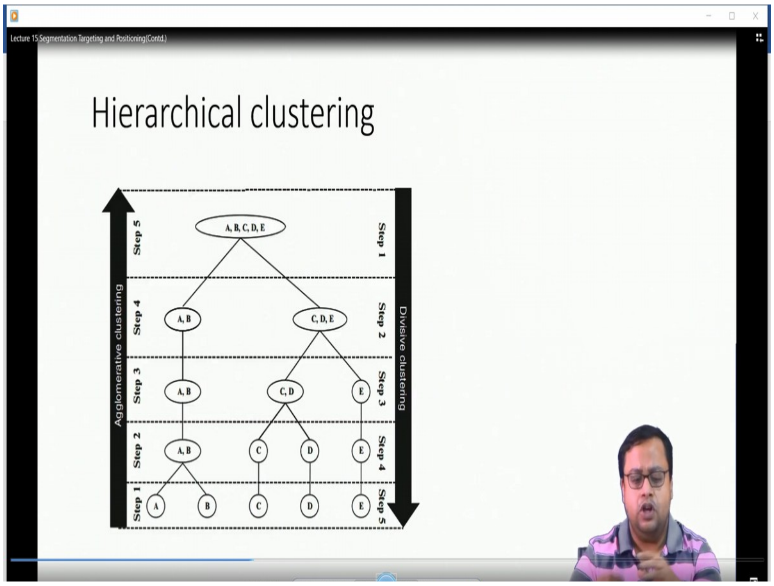
So what you do, you use age, gender, income, domicile and various other places, various other things as your 'x' variable, as your independent variable and 'y' variable is a categorical variable which is like….it can be two categories – segment A and segment B – for that you can directly use logistic regression, binomial logistic regression. If there are more number of segments, segments A, B, C, D, we use something called multinomial logistic regression. Multinomial means there are multiple names, binomial means there are two names, so we use multinomial logistic regression.

And we can also use Linear Discriminant Analysis which is another method by which we can probably predict that whether a customer will be in segment A, and segment B or segment C, based on the demographic factors. And based on that, I will create my product features and etcetera and I will target certain market rather than targeting all the customers. So this is something that is important.

So, here in today's class, what we will do is, we will not do the factor analysis part. Let us say, that we will have some data points and based on that, we will create segments and using those segments, we will also try to target those segments using the multinomial regression method. So I will also give certain inputs or certain information about how to, on the same dataset how to use LDA, linear discriminant analysis and what is the meaning of that in a different class.

So, today, we will talk about customer segmentation. So far these are the three steps that I have talked about, the four steps – hierarchical clustering, one of them is Ward's method and K-mean and model based. So we will majorly focus on the first three – Hierarchical, Ward's method and K-mean.

(Refer Slide Time 09:15)

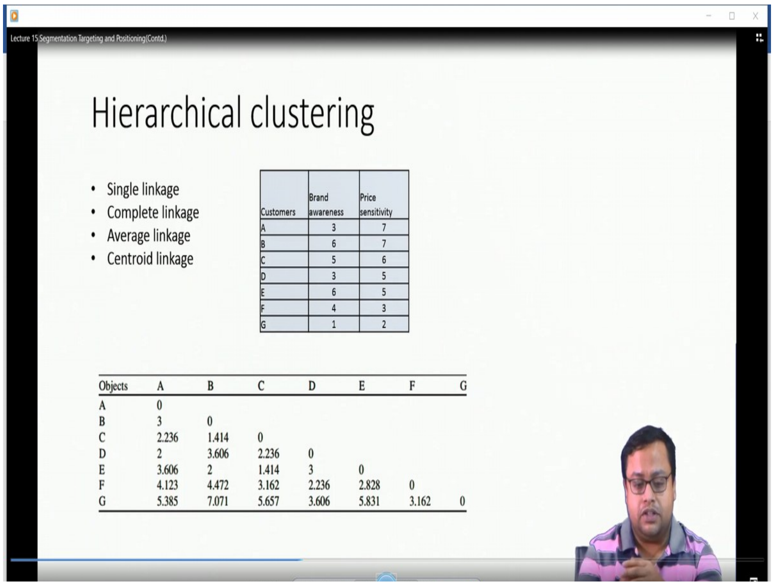


So hierarchical clustering looks like this. So let us say, you have five people – A, B, C, D and E and you start from below, so at the bottom if you see, that A, B, C, D, E, everybody are, A, B, C, D, E are all separate. So all the five people are in five separate segments. Then what is my job? My job is to join them, to create, to reduce the number of segments and to join them and to put them in the same segment. So I find out let us say A and B are very close. So I join A and B and in step two, I have four segments instead of five. A and B gets joined and C, D, E. Now for further analysis, I have to find the distance of C, D, E with the joint AB as a segment rather than A individually or B individually.

So next, I see probably C, D are the guys who should be joined. So I join C, D. In the third step, I have three segments and then again I join C, D and E and then I join A, B, C, D. So I slowly go on joining. So if I keep on joining, the maximum possible number of segments is all five but each of segment consist only one customer. So all five has only one customer each. So that is the maximum possible number of segments. What is the minimum possible number of segments? Yes, everybody is in the same segment. So there is only one segment available.

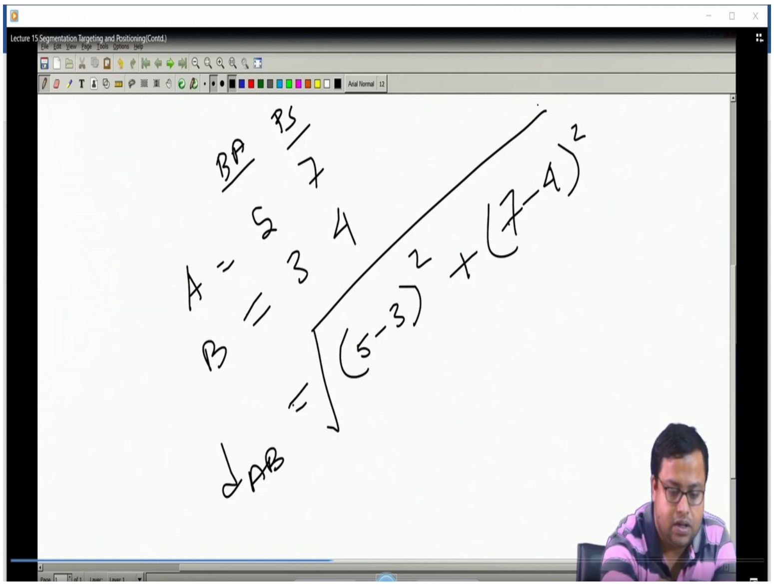
So you can do many things – you can either start from the bottom and go to the top, that we call as agglomerative clustering and I can also do the opposite one – I can start from the top and go to the bottom which we call as divisive clustering. Both of them is hierarchical clustering because they are creating an hierarchy of that. And how to do that?

(Refer Slide Time 11:11)



So there are several methods of finding out the distance between two groups or two people and etcetera. So let us say, A, B, C, D, E, F, G- I have in this case, seven customers and I have created two major variables which is brand awareness and price sensitivity and these are the scores given for brand awareness and price sensitivity. So what happens is that I find, I try to find out Euclidean distance between person A and person B. So what is an Euclidean distance?

(Refer Slide Time 11:54)

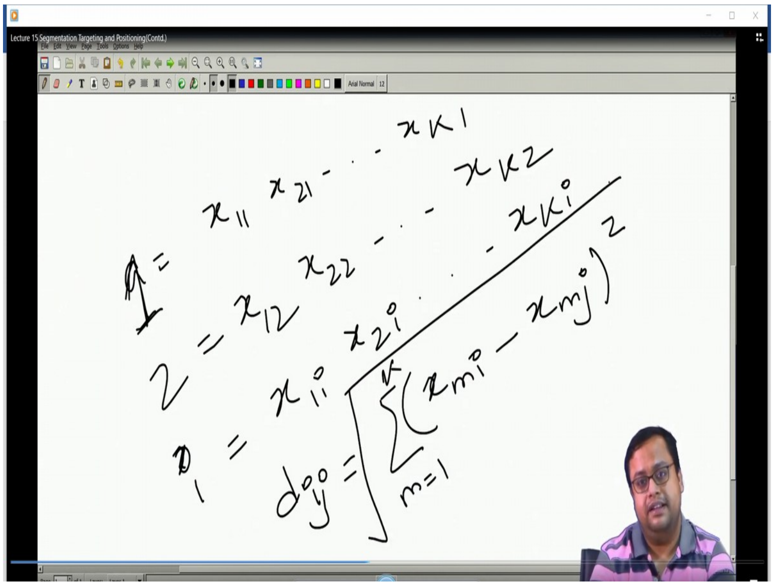


So the Euclidean distance formula will be, let us say, if A has five brand awareness and price sensitivity as 5 and 7 and for B, it is let us say 3 and 4, so the distance between A and B is

$d_{AB} = \sqrt{(5-3)^2+(7-4)^2}$

That is an Euclidean distance. So we can try to find out Euclidean distance. So what is the Euclidean distance formula? The Euclidean distance formula is something like this:

(Refer Slide Time 12:34)



So the formula is like this: that if A has x_1, x_2……..up to x_k number of characteristics, so person one, rather than A and person two has, so x_{11}, x_{21}…….x_{k1}, number of characteristics. Person 2 has x_{12}, x_{22}…………x_{k2} number of characteristics, then what is the person 'i'? The ith person has x_{1i}, x_{2i}…………….x_{ki}, number of characteristics.

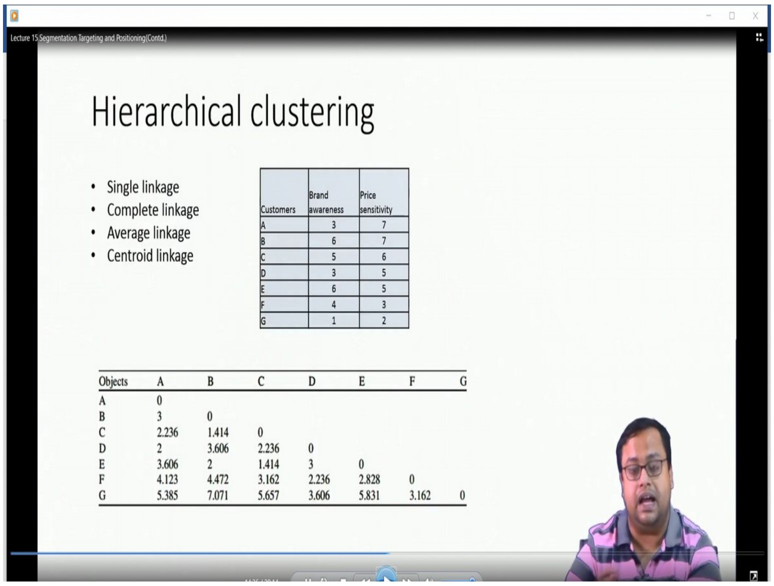
So let us say, there are two persons i and j. The distance between i and j is basically

$d_{ij} = \sqrt{\Sigma (x_{mi} - x_{mj})^2}$ [for m=1 to k]

So that can be the formula of Euclidean distance. So each of the two, same characteristics, you take it up – let us say brand awareness of A, brand awareness of B, you take it up, find out the distance, square it up. Then again you find out price sensitivity, it's of A, it's of B, do a subtraction, square it up.

And then let us say something else, let us say brand loyalty of A and brand loyalty of B, subtract it and square it up. Then you join them, add them up and take a square root. So that is how Euclidean distance is calculated. So, once we calculate the Euclidean distance, we can calculate actually 7C2 number of Euclidean distance. So each of the guys will have Euclidean distance with each of the people. So that is there.

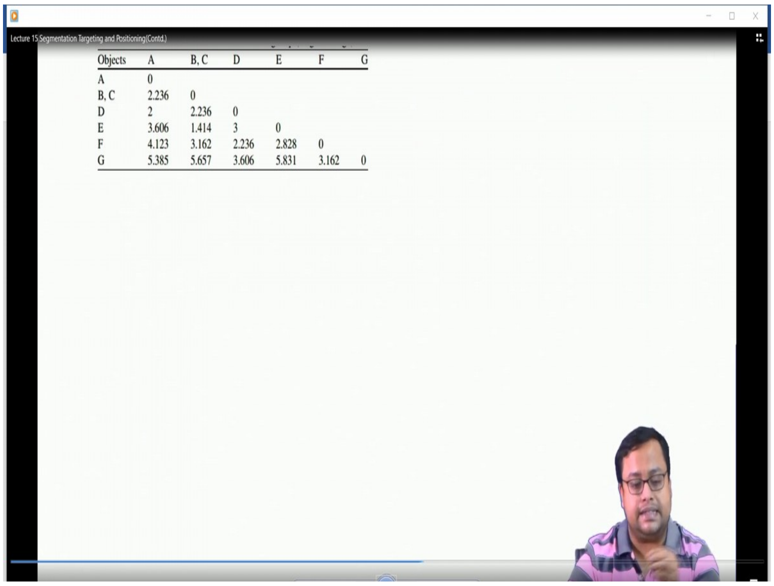
(Refer Slide Time 14:37)

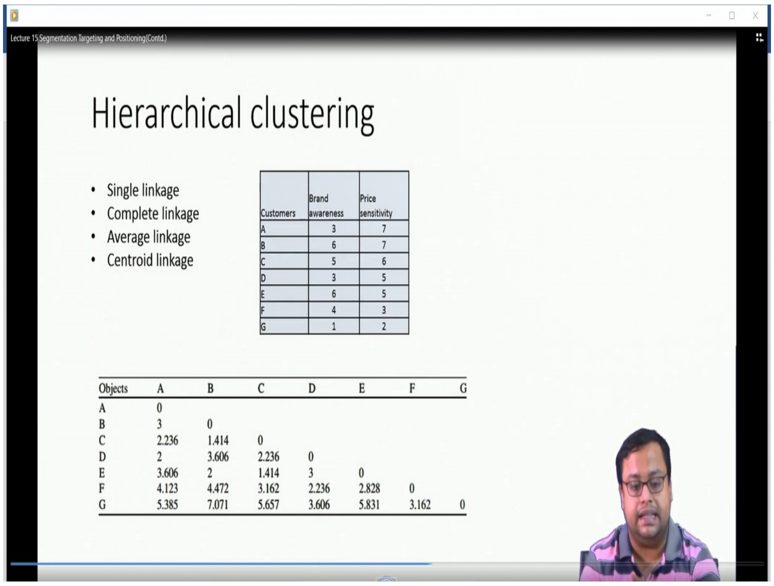


And then, what we do is, see, the diagonal elements are all zero and it is only a triangular matrix, that means, nothing above the top, why? Because whatever is above the top, above the diagonal is actually same to whatever is below the diagonal. So if the distance between A and B is 3, the distance between B and A is also 3, so that is why we only focus on any one part of the whole matrix. Now you see, which two guys are of least distance in this particular this thing. In the matrix that you can see here which one has the least distance?

So you can see that the least distance is coming up for C and E and for B and C. So C is the person who has the least distance with both B and E. So if that is the case, then what will you do? So if that is the case, then you can either join B and C or can join C and E. So any two guys you can join.

(Refer Slide Time 15:55)

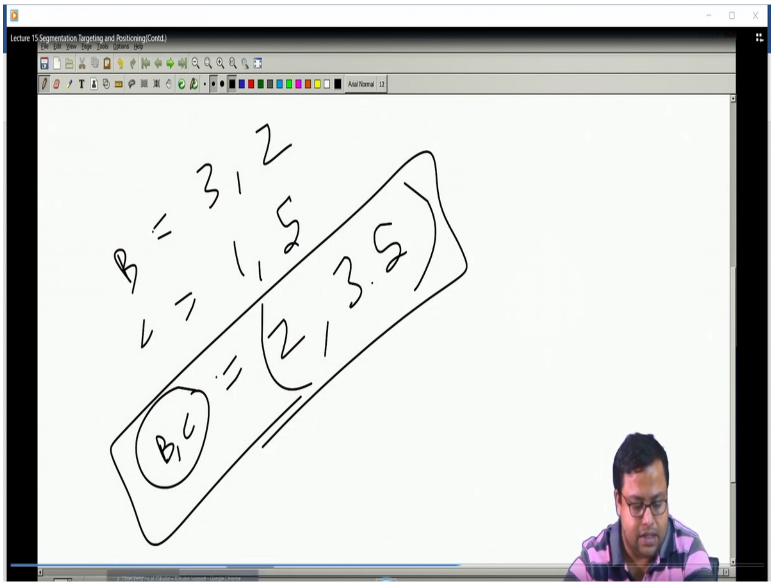




So I randomly started, I started by joining B and C. Now BC becomes one single segment. Now there are different ways to find out, how I can find out the distance from B and C to A, B and C to D, B and C to E and so on?….so which one?

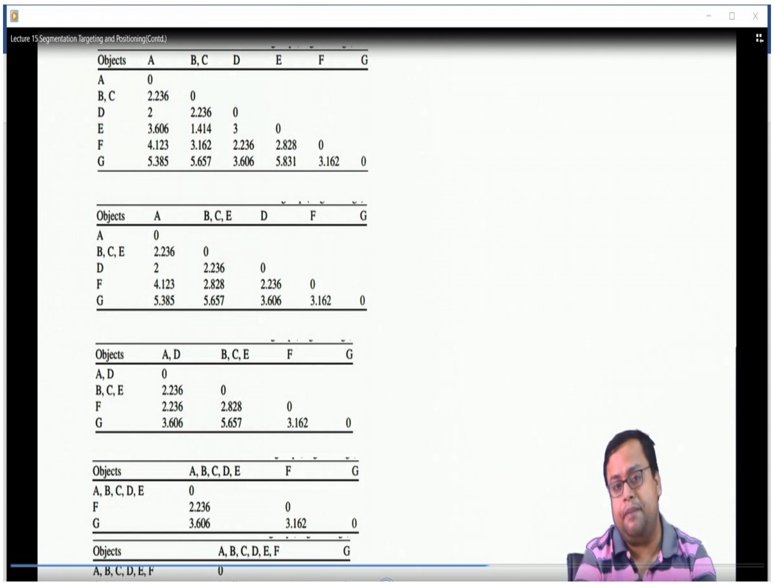
So that is how single linkage, complete linkage, average linkage and centroid linkage are different. So you either find out the average of these two and then find out the distance of that from each of the observation. So you find out the centroid or sometimes you find out the complete linkage. So, you take each of them distance from, so let us say BC is your segment, and in complete linkage, you might find out B's distance with everybody and C's distance with everybody and then take an average of that. So there are different ways of calculating the distance. The easiest way of calculating the distance is actually taking the mean.

(Refer Slide Time 17:00)



So I would suggest that if that is the case, if you have two guys, let us say B and C got joint, so B is let us say 3, 2 and this guy is 1, 5. Then BC when it got joint, the corresponding value becomes 2 which is the mean of 3, 1, 5 plus 2, 7 by 2 is 3.5. So you take this, the mean deal value, the average value of the two guys, then you go ahead and try to find out the distances. So you find out the distance and then the next distance which one is lowest?

(Refer Slide Time 17:29)



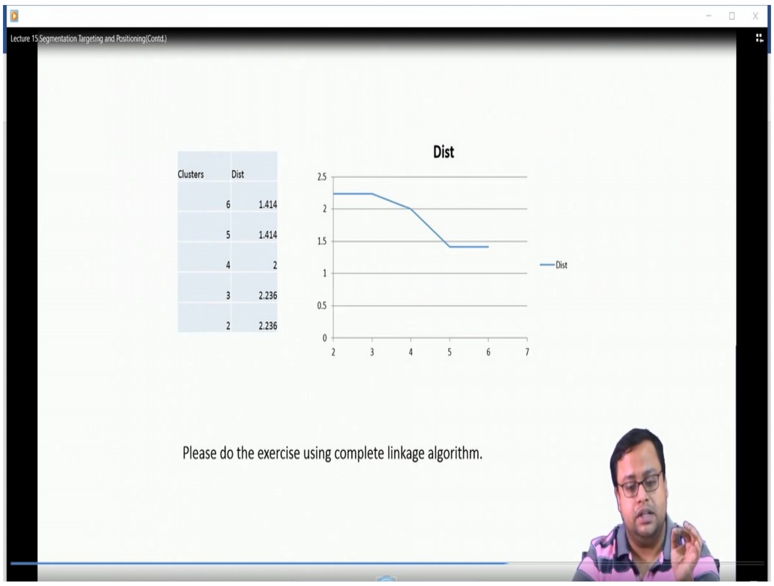
The lowest is BC with E. So here I think they have not taken the average. Here they have taken the single linkage which is the lowest distance. So BC and E. So BC and E is still

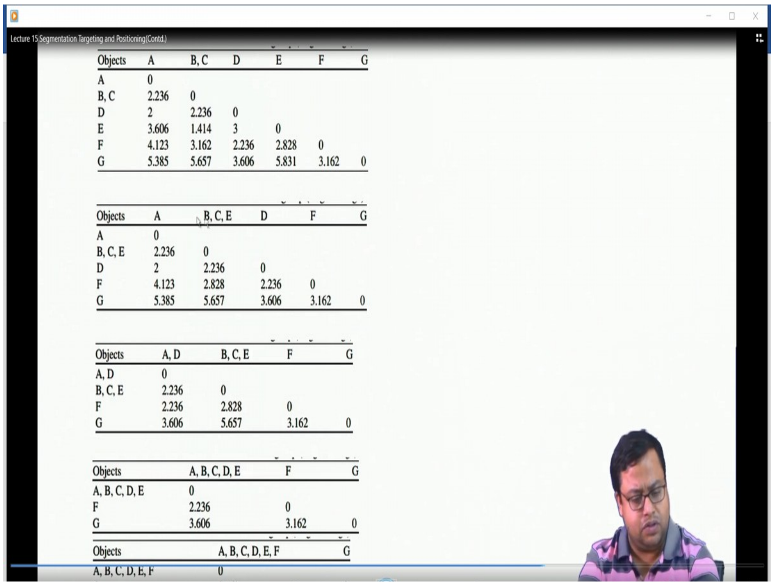
1.414. So then if I join them, it becomes A. So right now, you have how many? Five segments. Now I have five segments. I can still join, which one will I join? The lowest distance is A and D, so I join A and D. Now which one will I join? I will join either AD with A for AD with BC.

So let us say I join AD with BC because the size of the segment goes up by doing that and then what will I join? I will join ABCDE with F. So ABCDEF, I join. So I can keep on joining. Last one will be, everybody will be in the same segment.

Now this does not make any sense. If I keep everybody in separate segments, it does not make any sense because I cannot create products for each of the people individually and if I join everybody in one segment, that also does not make any sense because as we discussed that not everybody likes lukewarm milk, lukewarm tea. Some will like hot tea and some people will like cold tea and you have to find out which one you will target. You cannot actually create a lukewarm tea. So putting all of them in one segment does not matter. So then there has to be a point when I will stop. So should I stop at four segments, three segments, two segments? Where will I stop? That is something that is important.

(Refer Slide Time 19:15)



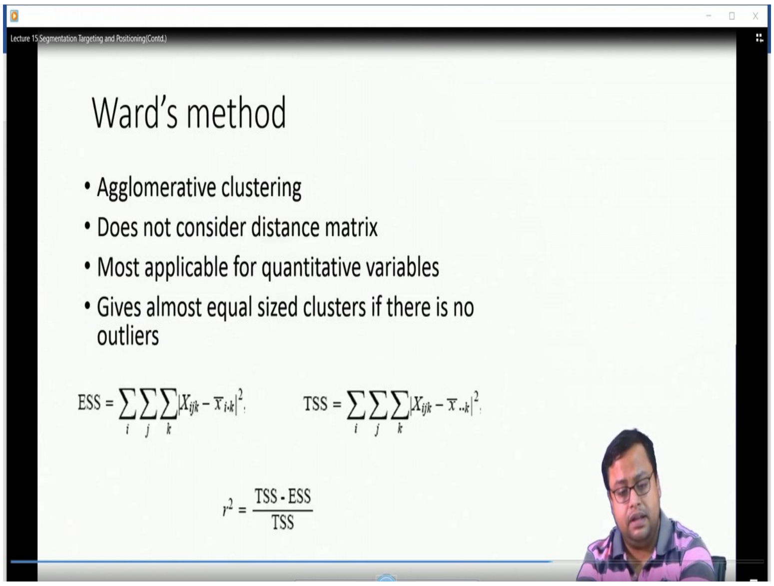


To do that, what we do is we plot. So we see that, when we started with from six clusters, from seven to six we jumped, initially everybody was in separate clusters or segments. When we jumped from seven to six, the distance covered was 1.414 and then when we jumped from six to five, again the distance covered was 1.414. The lowest distance was 1.414. And then when we covered from five to four, the lowest distance, the distance covered was 2. From four to three, it was 2.236 and from three to two, it was also 2.236.

So if I plot that, generally what we find is a kink. What is a kink? It’s an elbow like situation. What is this elbow like situation? Or always you will get such good looking elbow like situation. What is situation? That means that before that, if you actually further increase the number of segments, you do not get much information but before that the information, extra information that we are getting is much higher. So this you can assume as a measurement of information that you get and this is where the elbow suggests that after that, further breaking does not matter, does not make any sense. So in this case, we have found five as our classic segment, so we can actually choice this.

One – A, D, F, and G are in separate segments and B, C, E are in same segment. So that is how you do. Now this is called Hierarchical method. Here we have taken Euclidean distance as our measurement.

(Refer Slide Time 20:59)



Ward's method is Agglomerative clustering. So that means it is a hierarchical method and does not consider distance matrix. It does not consider distance matrix, not applicable for, sorry, it is mostly applicable for quantitative variables and gives almost equal sized clusters if there is no outliers. This is some of the characteristics of Ward's method which is a special type of hierarchical clustering. So here, they actually measure the 'r square'. How much is the, instead of distance, they measure r square and check that which one I will join that will actually improve the r square.

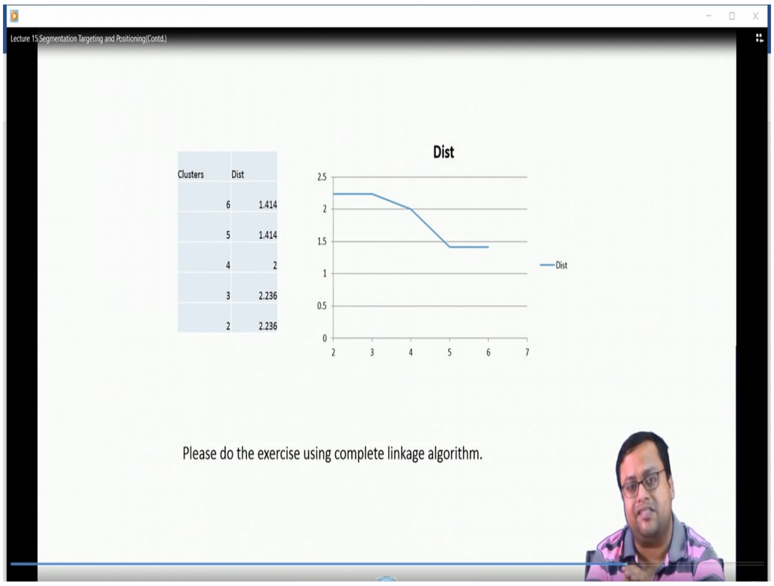
What is r square? R square is similar to what we have measured in case of regression. If you remember the multiple r^2, it is similar. So if you see that the first one, the ESS is the error that is still left. So X_{ijk}, what is the thing? The thing is that, so for all people who are in ith group, so ith group, so for all the people I find out the mean of that group and then divide it from that particular group's individual observations and then add them up. So that is one part which is the still error that is left. And TSS is, when I had no segment. When I had only one segment, when everybody was put in one segment, what was the situation? So that way, that is TSS.

And r square is just nothing but, it actually tries to find out whom I will put in which segment such that this r^2 maximizes. That means TSS minus ESS by TSS maximizes. It is similar to, so TSS minus ESS is what? Which is the part of the variance that has been explained. TSS is when there was no model, what was the error left? ESS was when there is some kind of clustering, what was the error left?

That means TSS minus ESS is what, the numerator of this r square is what, the percentage, or the amount of variance, not percentage, the amount of variance that has been explained by these clusters. So if I want to see in percentage form, that by TSS gives me the percentage of the variance that has been explained by these clusters and as r^2 goes high, we join people. So we try to find out whom to join such that the r^2 improves. So that is something that is also agglomerative clustering, that means, you start from all of them at separate segments and then you try to keep on joining.

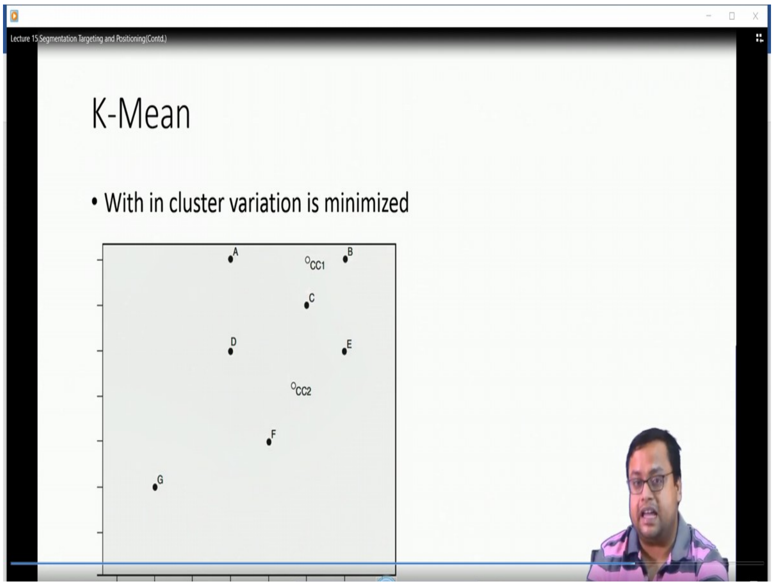
And here we measure; we actually plot this within sum of square as the measurement when we are measuring this.

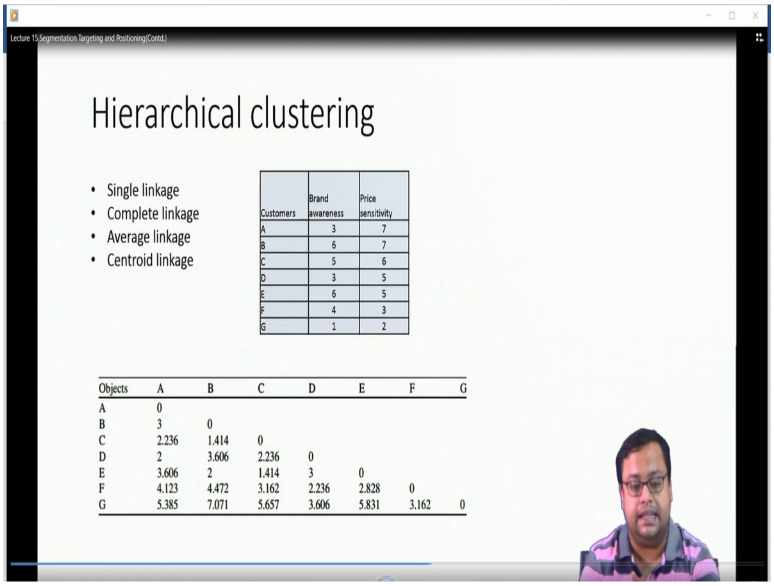
(Refer Slide Time 23:54)



We also plot Scree plot, we call it a scree plot here, so we also plot something like this but instead of distance covered, we measure within sum of square. That is Ward’s method.

(Refer Slide Time 24:10)





And the third method that is very easy probably is something called K-means method. So K-means method, so once we have decided through Ward's method or Hierarchical clustering method that how much will be the, how many segments you want and etcetera, you try to do something like, let us say you decided that I want to have three segments. Now you will try to use K-Mean to give a better application, sometimes K-Means works better than hierarchical clustering methods to find out who will fall in which segment, whether the segment is stable enough or not.

So what they will do is, let us say that I have two observations here also. So in a two dimensional plane, you plot the points. So for example, let's say, if you remember, A was

(3, 7), so brand awareness and price sensitivity is (3,7) , so what we did here is A here, you see, this is 0, so 1, 2, 3 and then this is 7. So 3,7 is A.

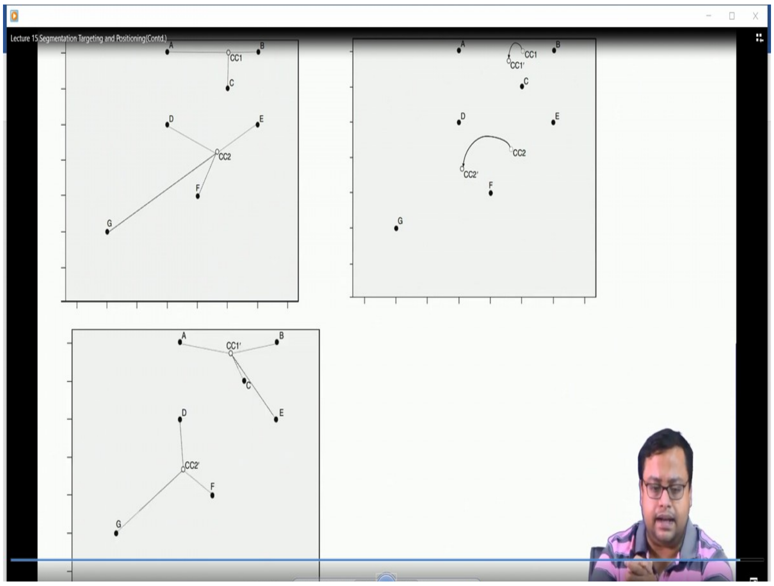
Similarly, for every point, I position them in this x, y plane. If it is a multi-dimensional observation, if there was let us say, instead of price awareness, I had some other variables, so the next method we will talk about is hierarchical clustering method. So sometimes in comparison to Ward's method or in comparison to the, sorry, hierarchical clustering method and Ward's method, sometimes the K-Means method works better. So the next topic is K-Means method.

What happens? Let us say, through Ward's method and Hierarchical clustering method, we have decided that we should have, through the Scree plot, we decided that we should have two segments or three segments. So what we do is then, we put the individual values like let us say, here we had A, B, C, D, E, F, G, there were seven observations and A was 3,7. If you check it, A was brand awareness, 3 and price sensitivity as 7. So you plot A, here you plat in x axis, you plotted A as brand awareness in the x axis and in the y axis, it is price sensitivity.

If there were more number of variables, then you create a multi-dimensional space where one axis is brand awareness, another axis is something else. So sometimes we cannot visualize a multi-dimensional space. For simplicity, that is why I have used a two dimensional plane. But in the multi-dimensional space, what you do is, you actually put these points and then you close your eyes and randomly put two points in the plane and see that how these two points, because you have two means, K-Means, that means two clusters will have two means.

If you have more clusters then more number of means, and how many clusters will come from Ward's method or the Hierarchical Clustering method. So you find out that, ok, I will have three or four clusters and you put those means and whoever has lesser Euclidean distance from the cluster mean, you just put that in that particular segment. So that is the first job.

(Refer Slide Time 27:29)



So what we do, I have positioned CC1 and CC2, two means and ok, so A, B and C are close to CC1 then to CC2 so I put them in cluster 1 and the rest D, E, F, G are close to CC2 then CC1, so I put them in cluster 2. But who says that this is something that, then we have to check how much this clusters are stable. So I shift CC1 and CC2 a little bit, delta x in each of the lines I would say, each of my dimensions.

So I shift CC1 and CC2 to new position, CC1 dash and CC2 dash. Now I see that whether the number of segments and I would say where the segment falls is same or not. And I see that all of a sudden, this E guy has come to CC1 instead of CC2. So that is something that I generally keep on repeating and I check that why it will become stable. The place where it becomes stable that is my final segmentation method. So if I keep on doing that, keep on doing that, at some point of time, I will reach stability and that will be my segments. Now a segment will be a good segment only when it is stable over all these three methods that we have discussed.

So we will continue on this particular video in the next part where we will do the coding with the dataset and we will use these three methods of clustering and then we will try to see that how targeting can be done. Thank you very much. I will see you in the next video.