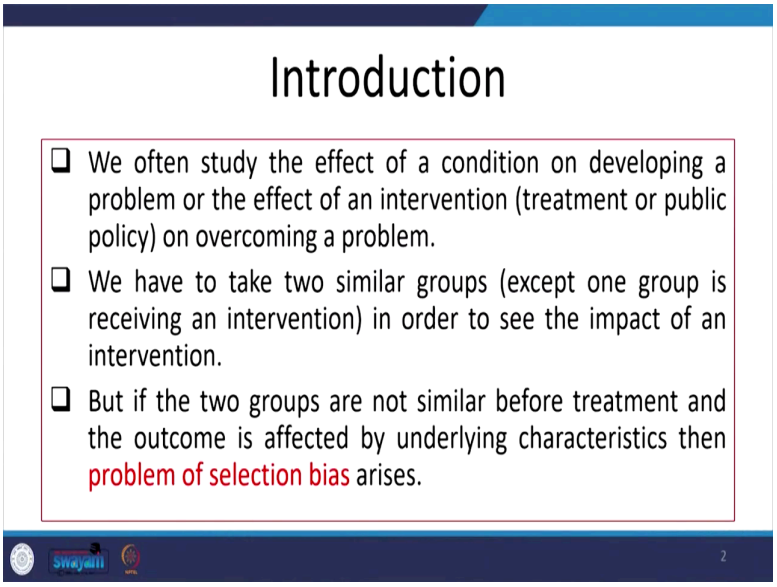
Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 37
Propensity Score Matching (PSM)

Welcome participants once again to the NPTEL MOOC module on Exploring Health Data. We are on the verge of the last week. We are we target ourselves to the understanding of Health Care Data and their programming, especially how to evaluate some of the health policies. In this regard, we have already discussed a lecture on clarifying policy evaluation and different techniques that are largely used in the previous lecture.

In this lecture, we will be emphasizing on what is called Propensity Score Matching and this is one of the designs, which is largely used both in the experimental case as well as in observational cases case studies. So, there are a number of decisions to be taken while going for propensity score matching that is in short called PSM. I am just going to introduce you all this is on how what are the genesis of PSM.

(Refer Slide Time: 01:34)



We often study the effects of a condition on developing a problem on the effect of an intervention. So, an intervention that is called treatment or public policy on overcoming a problem. Like, suppose we say any policy let it be in a portion of Iran that has impacted the

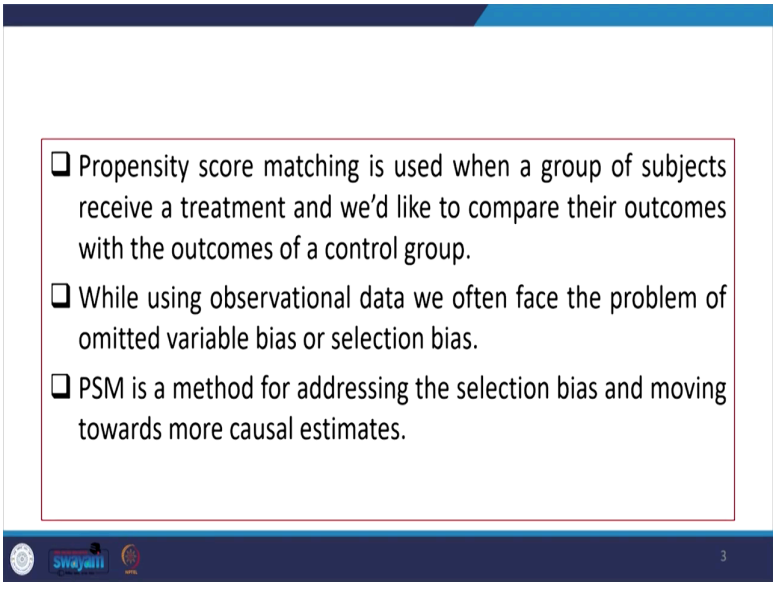
people due to any sort of intervention may be an insurance scheme, whether that has actually undermined the cost of expenditure of the persons or not.

So, you can do a field experiment or you can also go through the database, those are available through observational studies, and can identify a number of gaps within it and can suggest whether some of the policies are effective or not. So, we have to take two similar groups in this case that is except one group is receiving an intervention or treatment in order to see the impact of an intervention.

But if the two groups are not similar before treatment and the outcome is affected by underlining characteristics then problems of selection bias occurs. So, like if you do not equalize the starting two set of groups wherein one set of group you are having certain treatment whereas, you are not having treatment, but this group by definitions are not same by assumption they are not same or by starting point they are not same.

So, how do your treatment has actually caused certain differences or your policy interventions has intervention has caused certain differences, that is why if you are do not go with similar two groups. So, that may lead to selection bias. Propensity Score Matching is used when a group of subjects receive treatment and we would like to compare the outcomes with the outcomes of a control group.

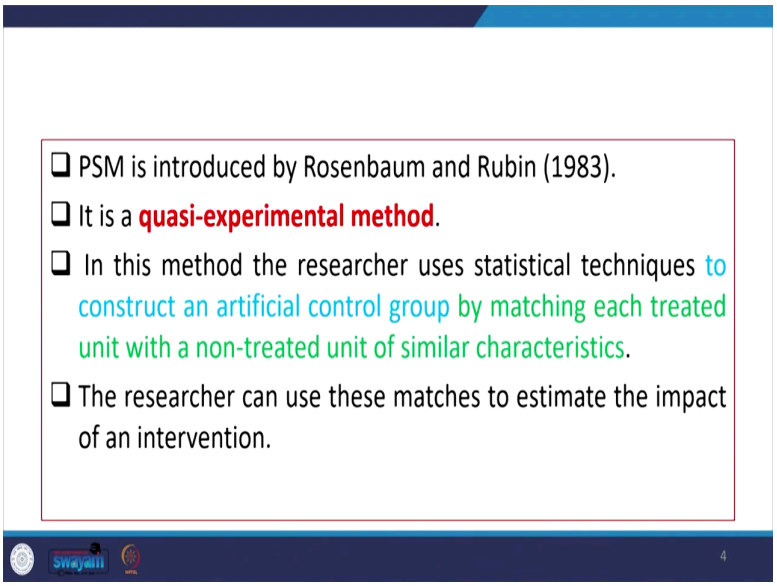
(Refer Slide Time: 03:32)



While using observational data we often face the problem of omitted variable bias or selection bias. So, either by the selection, we are omitting variables or there are some biases or by the variables are omitted so by some of the reason that may create biases in the model.

PSM is a method of addressing the selection bias and moving towards more causal estimates. It is not like the cross-sectional or least square method or the normal regression technique this is rather give you a certain causal relationship in the estimation.

(Refer Slide Time: 04:15)



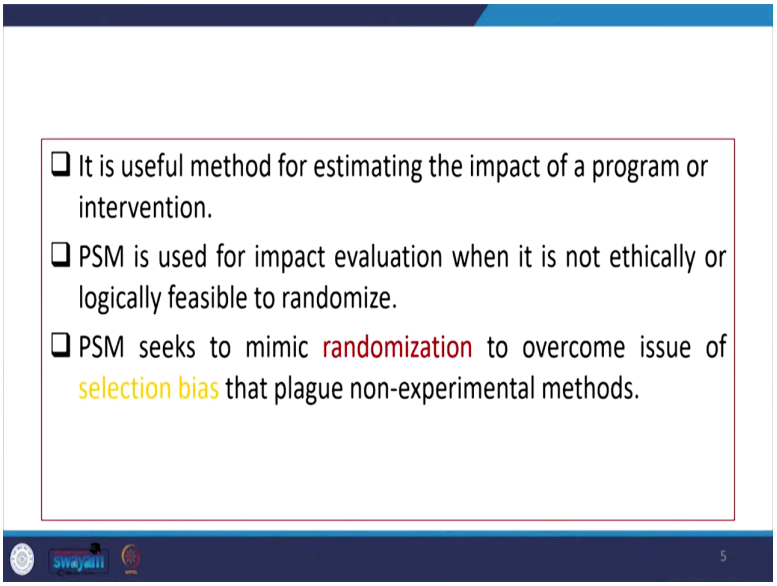
PSM is introduced by Rosenbaum and Rubin in 1983.This is also called quasi experimental method RCT is usually called the experimental design or method, but this is quite quasi not exactly tapping the impact of one with another one, there is no clear understanding of the cause and effect relationship.

In this method, the researcher uses statistical techniques to construct an artificial control group by matching each treated unit with a nontreated unit of similar characteristics. So, first of all an artificial control group is created. So, now what are the reasons what are the logic behind artificial the control group is not clearly defined, and there should be the non-treated group should have also the similar characteristics as the treated group.

So, the treated and non-treated groups the assumption since by assumption we say they are similar. So, anyone could be considered as the control group, and that is why we are saying any arbitrary one we are just speaking of to define this as the control group is in fact called

the artificial control group. The researcher can use these matches to estimate the impact of an intervention and it is useful it is one of the useful methods for estimating the impact of a program or intervention.

(Refer Slide Time: 05:45)

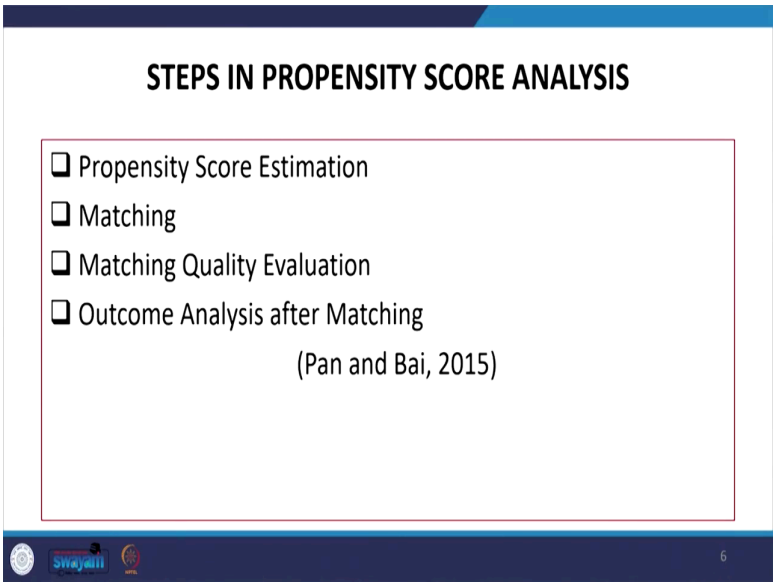


PSM is used for impact evaluation when it is not ethically or logically feasible to randomize. In reality, as we are going to discuss randomized control trial as well in our next class, we shall understand that randomization is really difficult. There are lots of possibilities of arbitration lots of possibility of a relationship.

So, of the set of data that is being adopted. So, if wherever you have certain doubts with randomization this is one of the plausible method one of the best method to be applied. That is why PSM is widely applied.

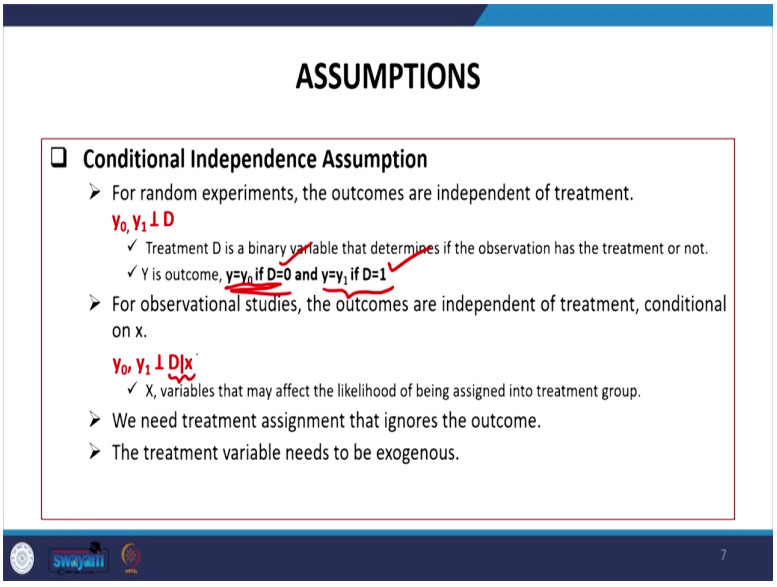
PSM is used for impact evaluation in because of these logical reasonings. PSM seeks to mimic randomization to overcome issue of selection bias that plague the non-experimental method. So, this is considered to be randomized because there is no difference between the two groups in that is how we can avoid the selection bias.

(Refer Slide Time: 07:06)



There are different steps involved in propensity score analysis, those steps are called Propensity Score Estimation Matching, Matching Quality Evaluation and outcome analysis after matching, once the matching result has come we need to have the comparison as well.

(Refer Slide Time: 07:27)



Let us talk about the assumptions of this PSM. The assumptions are like first one is called Conditional Independence Assumption. What does this mean? For random experiment, the outcomes are independent of treatment. Whereas, in case of observational studies the outcomes are independent of treatment conditional on the x.

So, when you are giving a certain treatment in the experimental model experimental design, your outcomes are independent of treatment. Whatever the outcome you are expected to derive are in fact independent. The, I mean your outcomes are independent of the treatment that is giving. So, a treatment that is being injected is not creating any sort of biasness to your outcome.

So, that does not mean the that any policy intervention is made in two areas. So, the policy intervention does not have a purpose of just motivating one area, just to show the result, just to show the outcome. So, that assumption has to be considered in the experimentation of PSM design. Similarly, in observational studies where surveys are conducted, it is conditional on the x.

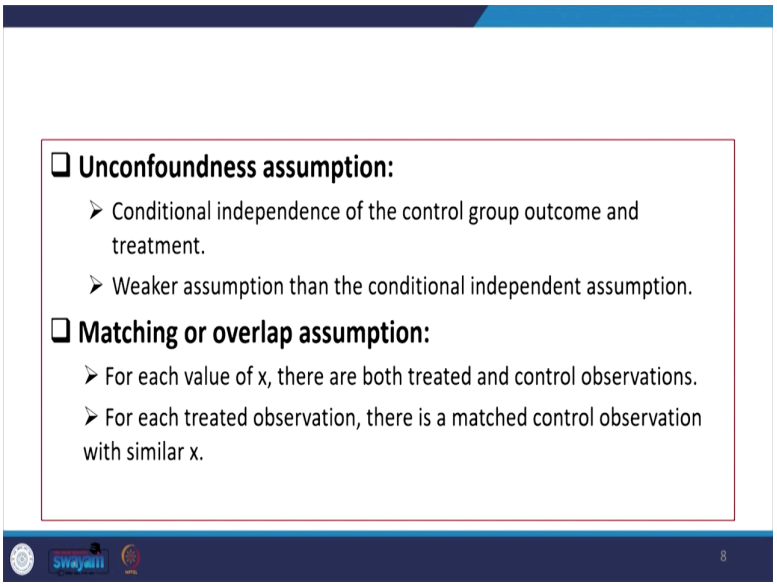
That x that is the control variables whatever we are taken, that those should not be also should be also independent of the kind of treatment that is taken, So, in the case of experimental design we know that the model is like y 0 y 1 these are the outcome in two categories, given the extent of treatment.

The treatment is a binary variable that determines if the observation has to be treatment or not. y is the outcome and y is equal to y naught if it is the base output which means, there was no treatment.

So, this is called actually no treatment and if it is the new output, that is due to the treatment. So, treatment is considered to be 1 here and here treatment is 0. In case of observational studies, we are conditioning here with the control variables, those treatments are subject to the control variables or but those are independent also of the where treatments were assigned.

We need treatment assignment that ignores the outcome this is what in the bottom line to be emphasized, the treatment variable needs to be exogenous. This is what is very essential it has to be exogenous.

(Refer Slide Time: 10:07)



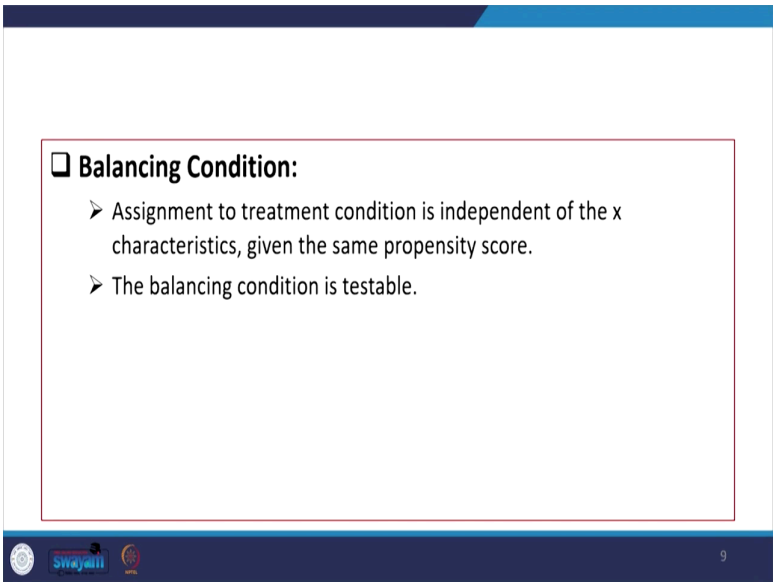
Some other assumptions are also emphasized like on confoundedness assumptions this says that, conditional independence of the control group outcome and treatment, So, basically says that conditional independence of the control group outcome and the treatment.

So, there should be independent-ness between the control group basically, in the next one which we have already said so this is how where independent of the condition of the control groups, So, these actually control groups and are also sometimes referred to as a confounded group.

So, the assumption is that there should be unconfounded in the treatment approach. Weaker assumption then the conditional independent assumption; so the so, basically unconfounded-ness is one is basically one of the weaker assumptions than that of the conditional independent assumption. Another assumption is related to the matching or overlap assumption.

Matching is discussed as for each value of x there are both treated and control observations. For each treated observation there is a match control observation with similar the control values,

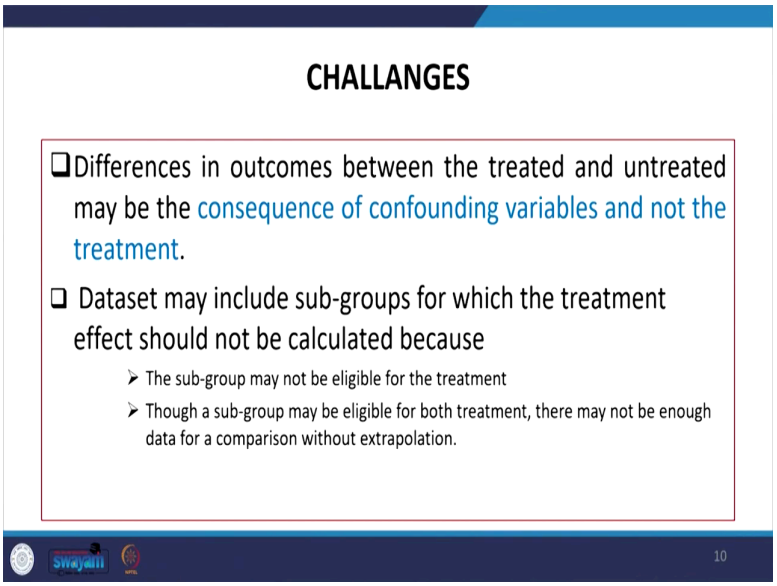
(Refer Slide Time: 11:31)



Then another one is called balancing condition assumption should have been also fulfilled. Assignment to treatment condition is independent of the x that we have been saying, x characteristics given the same propensity score. So, given the propensity score these treatment conditions should always be independent of the x.

The balancing condition is should be also testable. Though the balancing condition that has been followed in the model should have been testable, that is also the assumption. Then comes the challenges of this particular technique.

(Refer Slide Time: 12:10)

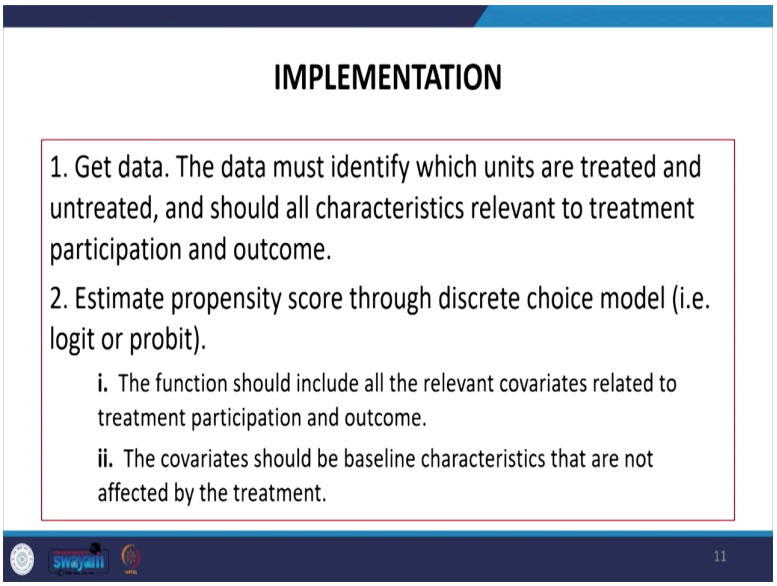


That is called differences in the outcome between the treated untreated may be the consequence of confounding variables and not the treatment. So, the challenge is that, though we are saying the treatment has resulted in certain differences in the outcome, it may be the due may be due to the confounding variables, the other confounding or the where the treatment is not given, on the cases though we have been saying several times.

But the variables may be some control variables are there which are nonother than the treatment variables. Those might have also caused the difference in the outcome. That is one of the challenges in the PSM method. The data set may include subgroups for which the treatment effects should not be calculated because the subgroup may not be eligible for the treatment. So, the subgroup is also important. All the treatment groups should not be carried on average, there might be sub groups that should be separated.

Actually, this is very difficult in the PSM approach to separate it out. Though a subgroup may be eligible for both treatments, there may not be enough data for comparison without extrapolation,

(Refer Slide Time: 13:32)

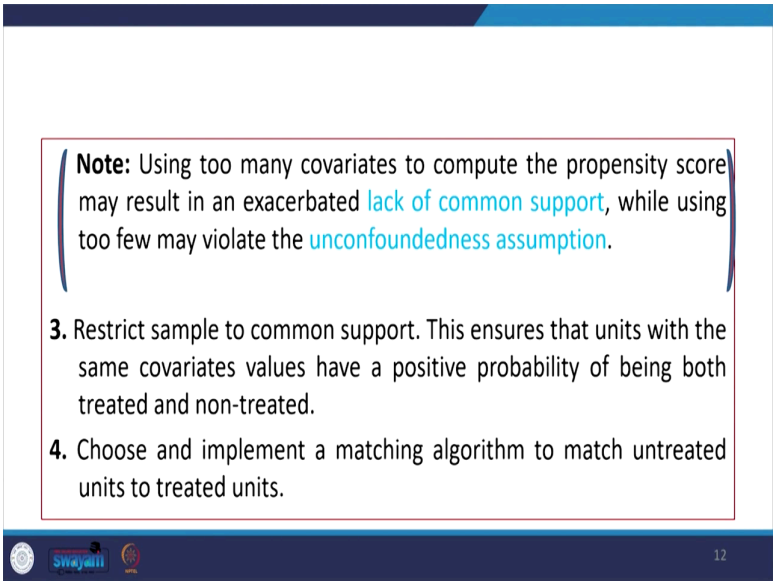


So, let us come for the discussion of the implementation of PSM, where we implement and how should we proceed. Now, we need to first get the data. The data must identify which units are treated and untreated and should all characteristics relevant to treatment participation and outcome.

Then the next we need to the second step is to estimate propensity score, through discrete choice model usually this is followed logistic or profit, or some high transitivity-based logic model logit model is also implied. So, a largest discrete choice model where logit or probabilities is considered for analysis. In this case the function should include all the relevant covariates related to treatment participation and outcome.

The covariates would be baseline characteristics that are not affected by the treatment, the covariate that we have been trying to refer to should be actually the baseline characteristics that are not affected by the treatment. So, treatment every time we have said that should be completely independent,

(Refer Slide Time: 14:49)



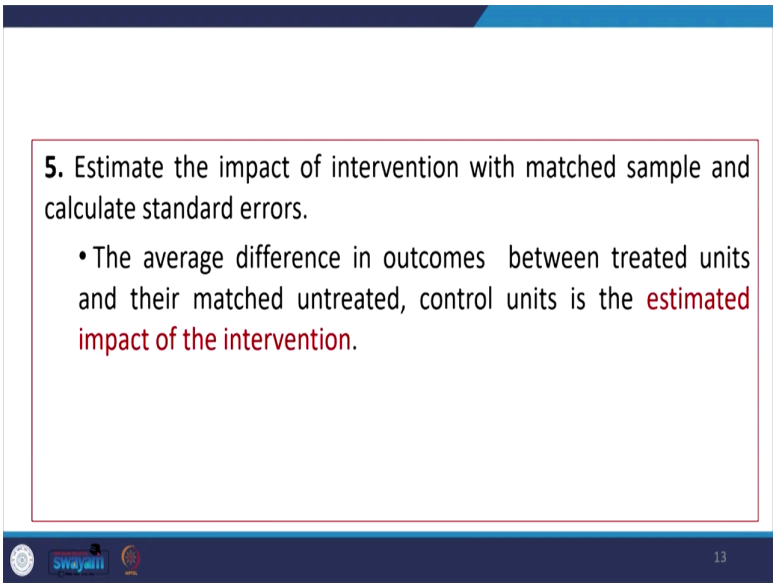
So, using too many covariates to compute the propensity score may result in an exacerbated lack of common support, while using too few may violate the unconfounded-ness assumption. So, confoundedness and the unconfounded assumption that we have already discussed. So, if you are including too less what covariates or the variables other than the treatment variables.

If you are taking too few then the unconfounded-ness assumption that is, how the confounding variables should be independent to that of the treatment is very difficult to understand. If there are too many then there might be a lack of support from the confounding variables.

So, there might be some problems we need to take a note on this particular regard. So, too many and too few though there are no parameter though extreme values defined, how many it should be so we cannot have a decision. Just based on our experimentation we can derive the circumference of these limits. Coming to next step of implementation is to restrict the sample to a common support.

This ensures that units with the values of the same covariates have a positive probability of being both treated and non-treated. Choose and implement a matching algorithm to match untreated units to treated units.

(Refer Slide Time: 16:25)



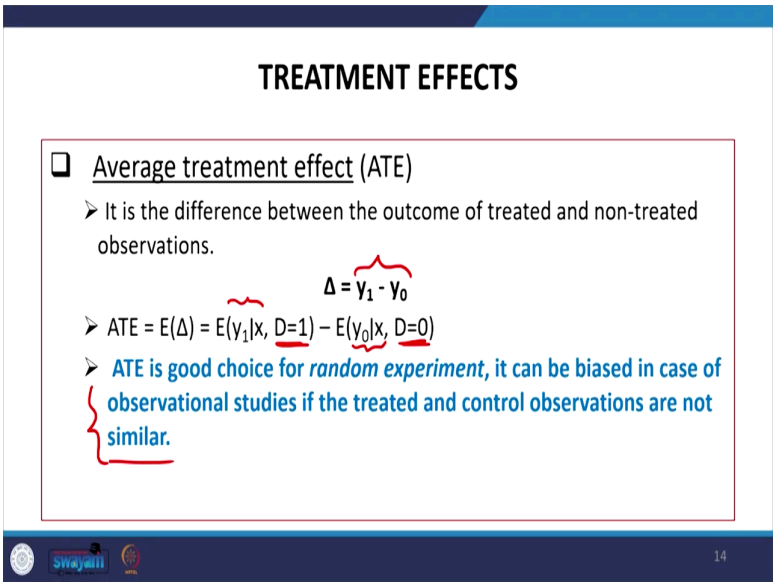
The last fifth one is on estimating the impact of intervention with a match sample and calculate standard errors accordingly. The average difference in outcomes between treated with their matched and untreated control units is the estimated impact of the intervention.

Basically, once you calculate the average difference in the outcome, between the treated one as compared to the unmatched but almost with the similar characteristics, while we make the difference the estimated impact at the value is nothing but is called the impact of that intervention through PSM.

Let us come to the understanding of the numerical values which we usually go through is called ATE and ATET those are also called treatment effects. One is estimated at the population level one is estimated at the core treated sample level.

So, when we are discussing about our entire population while we are discussing PSM those entire population the treatment average treatment effect, we calculate we can compare the control group and that of the treatment group which is what is called ATE. Whereas, within the treatment group of or the sampled population where intervention is made we can compare the implication the interventions,

(Refer Slide Time: 18:03)



So, let us understand this numerically, I will also experiment with the data with the technique and I am quite sure you will enjoy working with it. And there are so many other experiments that you can easily do it with this basic understanding. We are not going into the depth of or the complete applications of PSM because there are so many approaches followed by different authors.

So, just the guidelines we are giving or the baseline indicators or estimation we are explaining rest you can experiment and find it out. Coming to the average treatment effect in short called ATE. It is the difference between the outcome of treated and non-treated observations, that is basically the change between y 1and y naught this is what we already said.

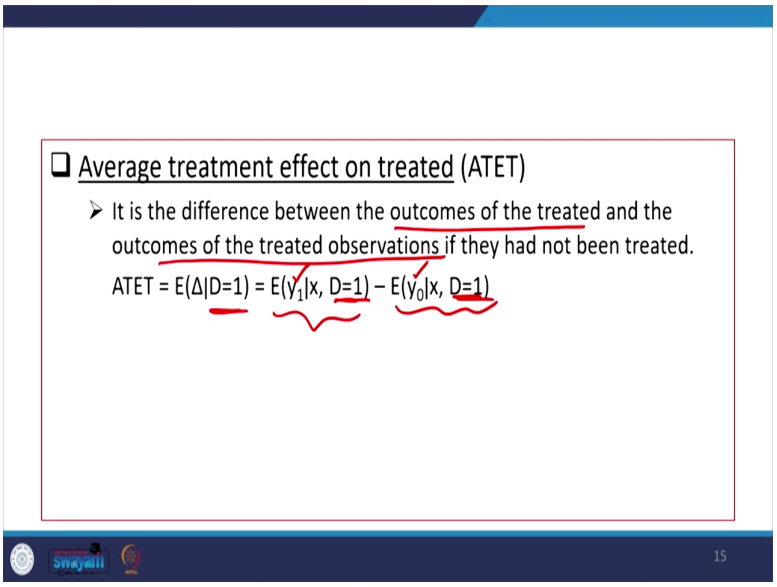
So, y 1 is the treatment group and y 0 is the control group or non-treatment group. So, that ATE is all about estimated value of that change is y 1 given the control variables, where treatment is given minus the expected value of y 0 given x where no treatment is given.

So, ATE is in fact going to give a choice about the random experiment and it can be biased in case of observational studies if the treated and control observations are not similar.

This is what we have been emphasizing several times ATE is good choice for random experiments. It can be biased in case of observational studies if and only if treated and controlled observations are not similar.

So, the basic structure of this ATE is that though there should be independent-ness between treated and controlled observations, similarly treated as compared to the confounding variables. So, in both cases if this is not similar then it might be problematic. Otherwise, ATE is a very good choice for random experiments.

(Refer Slide Time: 20:04)



Now, we are coming to the explanation of ATET. ATET is basically the Average Treatment Effect on the treated group. Within the treated group we wanted to find out the change, change due to the treatment. It is the difference between the outcomes of the treated and outcomes of the treated observations if they had not been treated. So, basically, when we think of some treatment, think of the treatment of certain intervention, it they start with no treatment actually.

So, no treatment and after treatment, does not mean they are the control group. Control groups are completely separated. The same treatment group what really happened when we

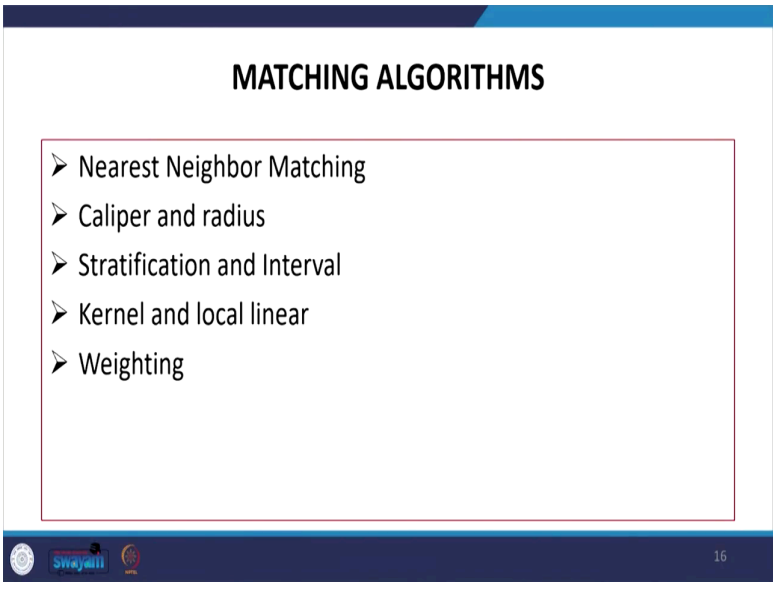
give treatment to those start to them from the starting point. That is basically called the sample, the final sample where we are targeted for the treatment.

So, the treatment group is actually compared between these two time periods, from the starting till the final outcome. Why we are trying to estimate because we wanted to understand whether the treatment group is actually different than that of the control group,

And or in the population sample how the result los like and in this particular treatment group within the treatment group how it los like, we can have a clear comparison for better understanding. So, ATET is basically a change in the treatment, a change in the value of a program given the fact that we are trying to estimate for that treatment group only where d is equal to 1.

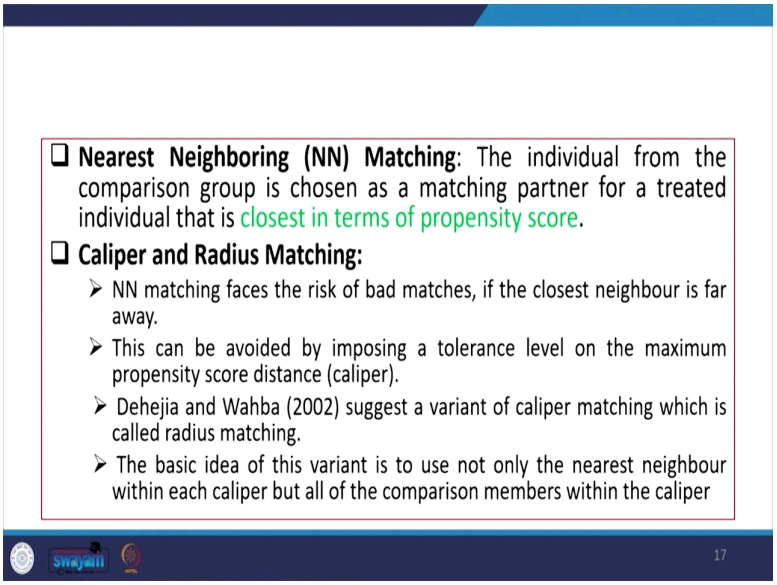
So, this is basically differentiating these as compared to this, but the difference is that in both the case it is equal to 1, but in this case, it is outcome 1 and this is outcome 0. So, without treatment for those treated groups before outcome and after the outcome. So, when we do that we get the ATET impacts.

(Refer Slide Time: 22:17)



So, some matching algorithms are also made one is called Nearest Neighbor Matching, then Caliper and radius Matching, then Stratification and Interval Matching, then Kernel and local linear method, then another is called Weighting method or algorithm.

(Refer Slide Time: 22:31)



So, those are also done. We are just giving certain ideas of each of the components here. So, Nearest Neighbor Matching is also in sort of called NN matching. The individuals from the comparison group is chosen as matching partner for a treated individual, that is closest in terms of the propensity score. When like in the command itself we will also show how whether it is choosing 1 is to 1 or 1 is to 2 or 3 matching.

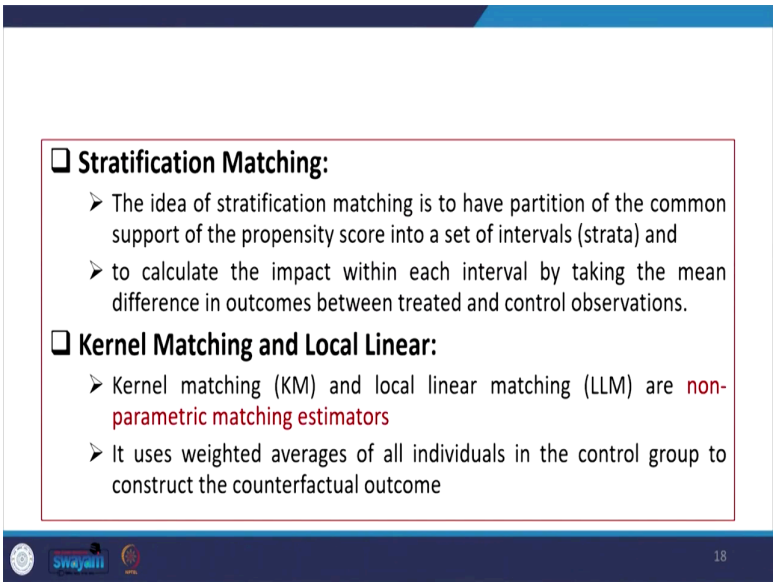
So, whether we are matching with the nearest neighbor or not is basically discussed with the NN matching approach. Another is called Caliper and radius Matching. NN matching faces the risk of bad matches, if the closest neighborhood or neighbor, is far away.

If there is no close value or close neighbor for the matching, then basically it is going to keep you bad results or bad matching in that case Caliper and radius Matching is more applied. This can be avoided by imposing a tolerance level on the maximum propensity score distance. When there is a distance between the neighbor's closest neighbor, in that case, you can talk take a tolerance level that will be helpful for matching that was suggested by Caliper.

The Dehejia and Wahba suggest a variety of Caliper Matching which is called Radius Matching. The basic idea of this variety is to use not only the nearest neighbor within each caliper but all of the comparison members within the caliper.

So, even in some studies we suggest that the caliper value should be less than 0.6 and the number of samples that were that is taken should not be less than 200 to run the PSM method, those aspects we are not emphasizing much that you can explore in your studies.

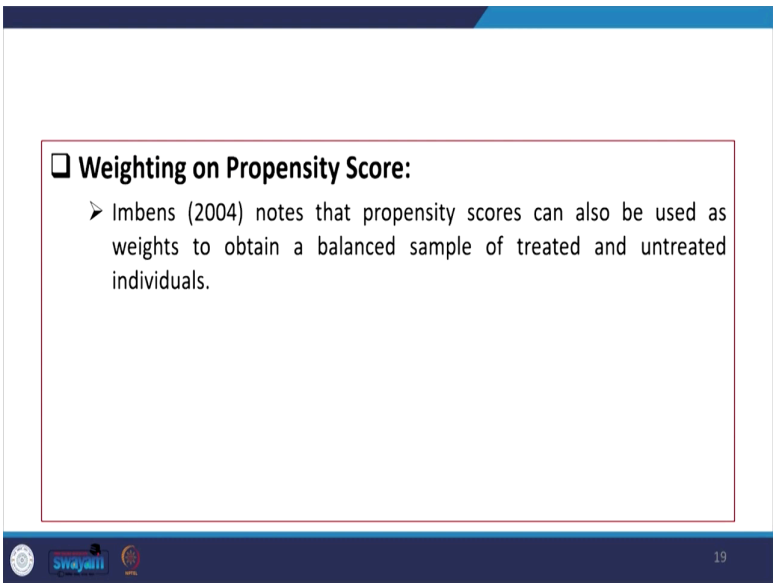
(Refer Slide Time: 24:40)



The next one is called Stratification Matching. The idea of Stratification Matching is to have a partition of the common support of the propensity score into a set of intervals. They are called strata and to calculate the impact within each interval by taking the mean difference in outcomes, between treated and control observations. The next one is called Kernel Matching and Local Linear. Kernel Matching and Local Linear Matching are nonparametric matching estimators,

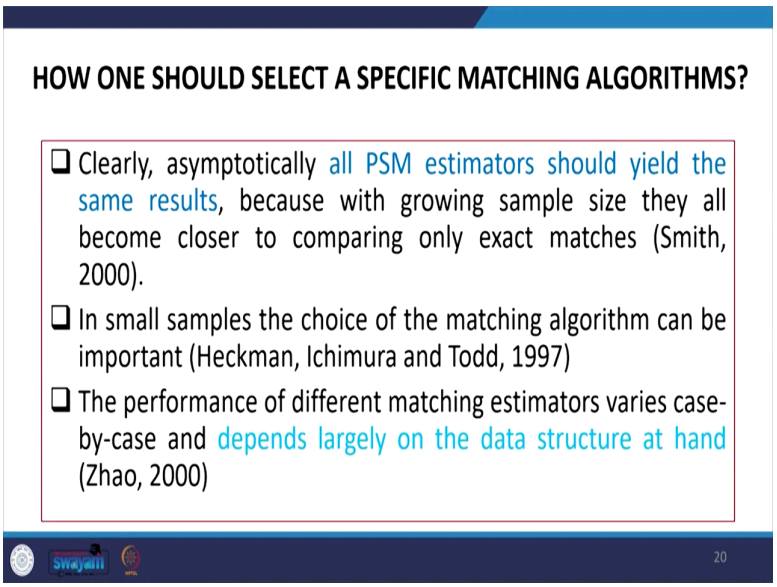
There are since they are actually having no parameters estimated. So, the only comparison is made. It uses weighted averages of all individuals in the control group to construct the counterfactual outcomes.

(Refer Slide Time: 25:23)



The last one is called weighting on the propensity score. Imbens to 2004 notes that propensity score can also be used as weights to obtain a balanced sample of treated and untreated individuals. So, the sample that we have been discussing may be on similar to this propensity score those can be balanced out,

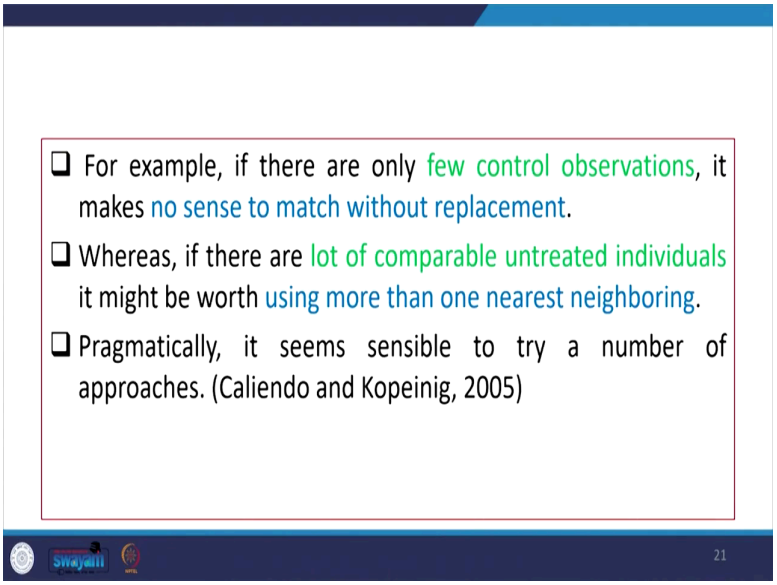
(Refer Slide Time: 25:50)



How one should select specific matching algorithms that is also important. Clearly asymptotically all PSM estimators should yield the same results. Because with the growing sample size they all become closer to comparing only exact matches as suggested by Smith's 2000 paper. In small sample, the choice of the matching algorithm can be important,

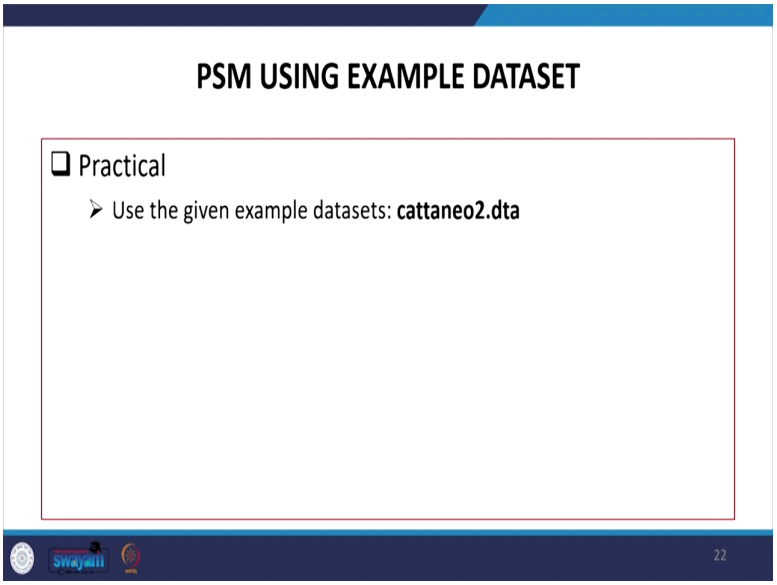
Choice of matching algorithm is more important where in case of large sample that is not so important. The performance of different matching estimators varies case by case and depends largely on the data structure at hand we have. For example, if there are only few control observations it makes no sense to match without replacement,

(Refer Slide Time: 26:38)



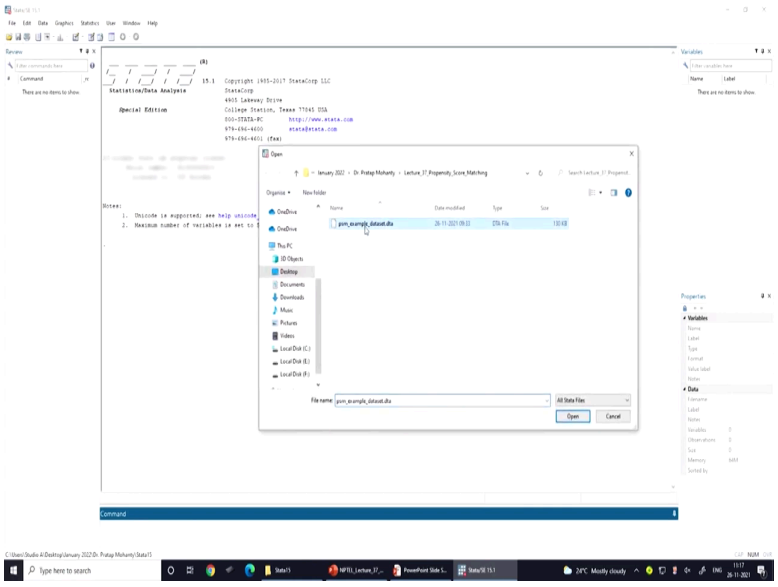
Whereas if there are a lot of comparable untreated individuals it might be worth using more than one nearest neighboring, pragmatically it seems sensible to try a number of approaches so that we can find out the robustness of the method.

(Refer Slide Time: 27:01)

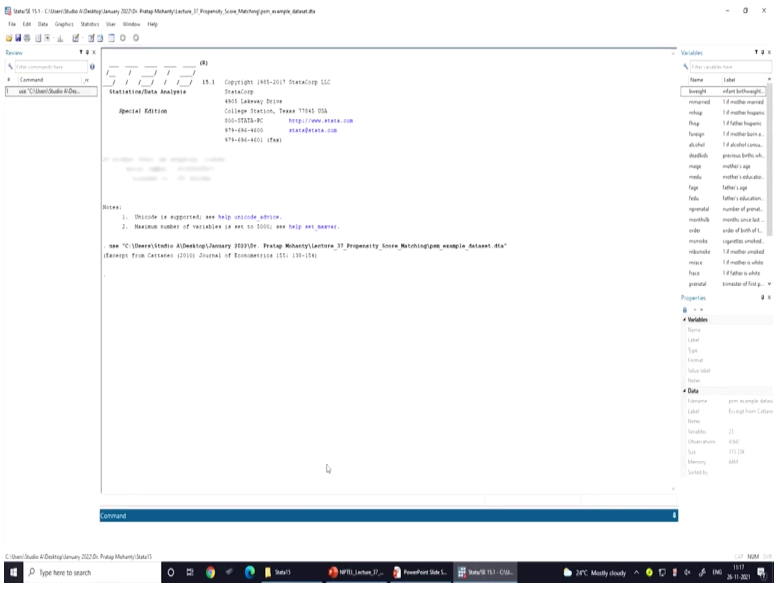


Now, we are going to discuss our practical approach to understanding the values of PSM, how to interpret it, we will use a sample data set. The sample data set is taken from the stata data set. Stata-defined data set that you can also download, the link we have already given over here. The data set we are also keeping on the portal; you can easily get it.

(Refer Slide Time: 27:34)



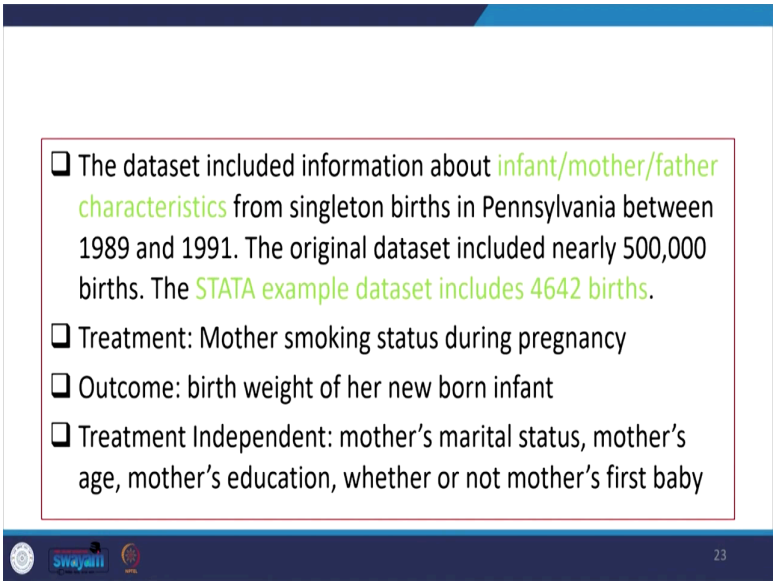
(Refer Slide Time: 27:46)



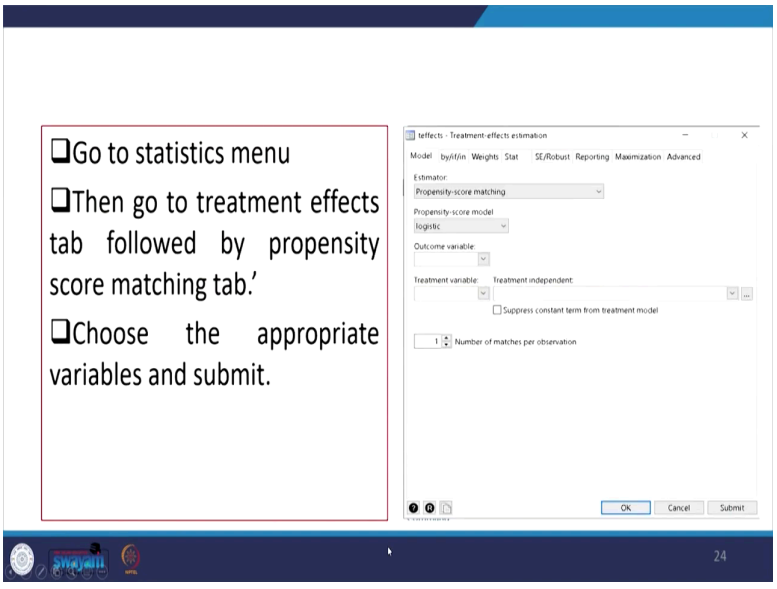
So, let us open the stata over here, and now on the screen we are going to open the data set. So, we are going to open the folder. The practical data set is PSM. Yes, we have opened this

data which we have shown to you. Now, we are going to run one thing that is going to be very useful for all of you is this.

(Refer Slide Time: 27:57)

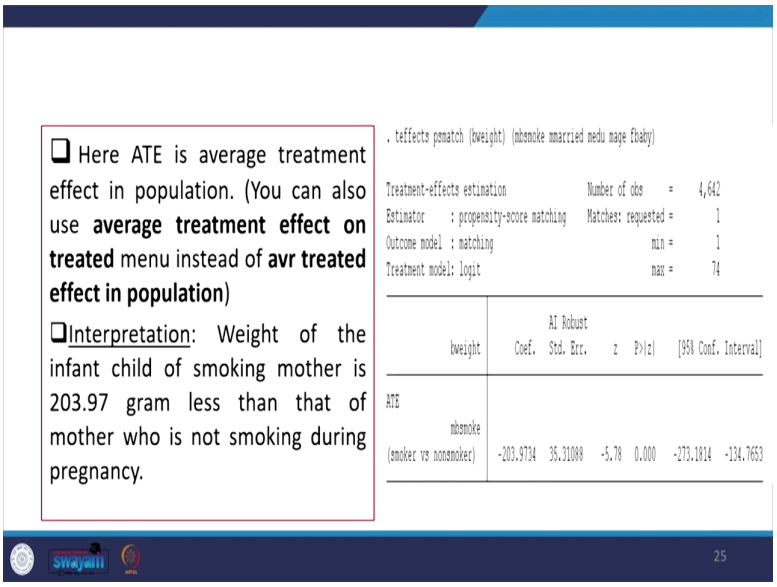


(Refer Slide Time: 27:59)



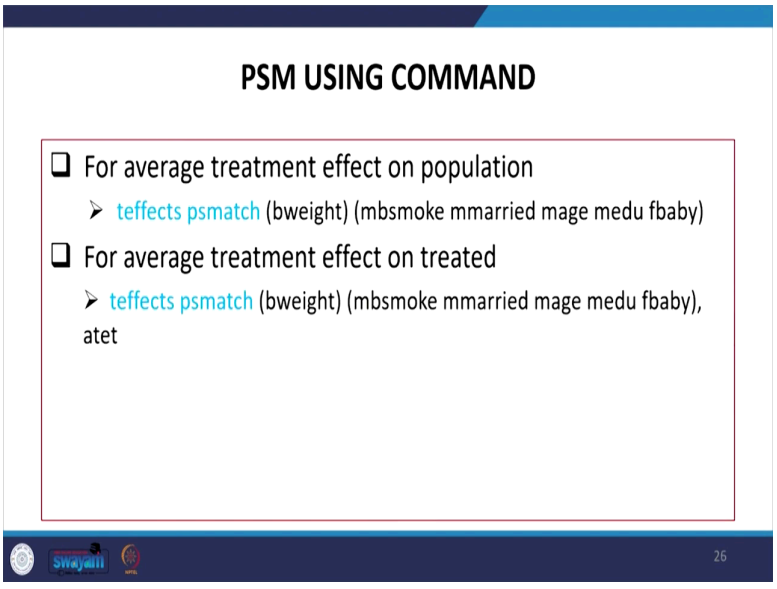
I am just going to give you some results over here first of all there are two approaches we will follow.

(Refer Slide Time: 28:09)



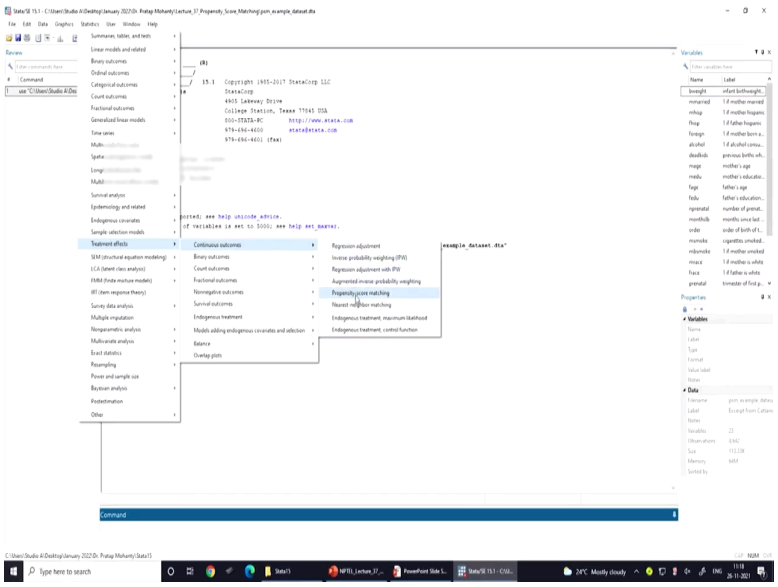
Let us go by the first approach, then we will come back like through click-based approach, we can do it then this kind of result we are going to find.

(Refer Slide Time: 28:13)



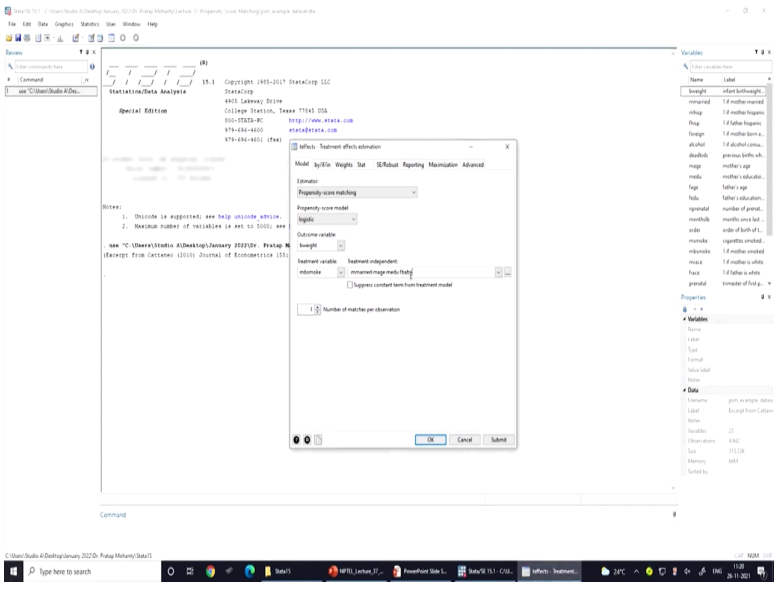
And then, we will interpret, and finally, we will come up with the command, So, let us go by this approach first. So, I am just going to open this on the screen.

(Refer Slide Time: 28:29)



So, we will go by statistic go to the statistics. Then treatment effects then we will click on the continuous outcome then on the propensity score matching.

(Refer Slide Time: 28:40)



From here we will take the data to estimate first of all here we will understand that I will just go to the data first. So, here we will show it like a couple of things. So, we are going to show it. The data set included information about infant or mother and father characteristics from singleton birth in Pennsylvania between 1989 to 1991. The original data set included nearly 5 lakh births. This data example data set includes 4642 births.

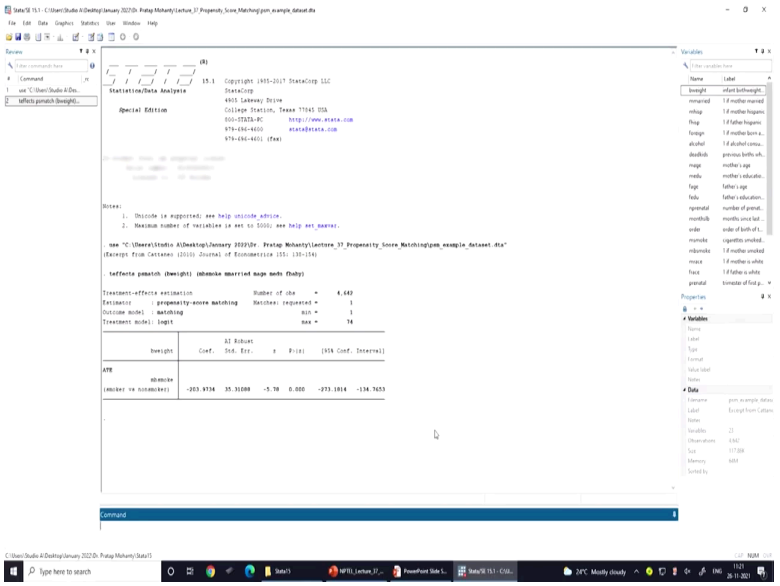
The treatment here is considering mothers' smoking status during pregnancy, and what this has caused the impact on the birth weight of our newborn infant. So, that would be basically our outcome variable, then treatment independent variables are like basically the covariates here for us are mothers' marital status. These are also called control variables.

Marital, mother's marital status, mother's age, mother's marital status, mother's age, mother's education; these three, mother's marital status, mother age, and mother's education and another variable we are considering whether the mothers have the first baby or not, So, regarding the baby than the mother's age, education, then marital status. So, these four we are going to use it for our calculation.

So, it is here on the screen. So, we will once again go to the point. So, we will outcome variable as I already mentioned. It is the birth weight of the baby, then the treatment variable is our smoking status of the mother, the pregnant women and so four variables we are going to consider covariate that is married whether the mother is married, or then age then education then another one is on that is called whether the mother is or had have had already a baby with them.

So, with these four with a number of matches per observation and it is matching with other values. So, if you select one only at this moment and we wanted to find out the result.

(Refer Slide Time: 31:21)



So, it is here. With, we get the result on the screen. So, this is going to give us couple of interesting interpretations. Now, the birth weight we wanted to find out, birth weight of the baby. Now, our treatment is like whether the smoking habits or smoking during pregnancy has impacted baby's weight or not. Now, we have seen that we can see here that it is actually impacting negatively.

So, score here the coefficient is going on the score the value by certain fall as 203 gram fall in the baby's weight its negatively and it is significantly related. And the method it is considered as logit method logit, Logit based on the distribution of the data by default you generally take logit,

But we can change when the distribution, we can change it to probit as well. Now, it has actually the model is significant and the birth weight of the baby is actually drastically reduced by 203 grams, due to smoking habit, alright. So, that is important and we have already discussed it. Now, I will clarify some other aspects.

So, this is what we have guided. You have to go to the statistics menu then accordingly you can choose and give the entries very important is our outcome variables, then treatment variables then treatment independent these the stata calls as treatment independent or for us we said control variable these are also called treatment independence.

So, these values should be independent to the treatment values treatment variables, then with the summation, we will get the result like this. Here the ATE is an average treatment effect, average treatment effect, and treatment effect in the population. We are actually comparing the population rate control with that of the treatment group, the groups are actually different and they are independent. You can also use average treatment effect on the treated menu instead of average treated effect in the population.

So, the interpretation as I already said it will be baby's weight will be reduced by 203.97 gram, this is what is explained as compared to the baby who are the mother who are not smiling, So, you can also take the note of a number of observations. It usually a number of observations should be higher in the PSM method.

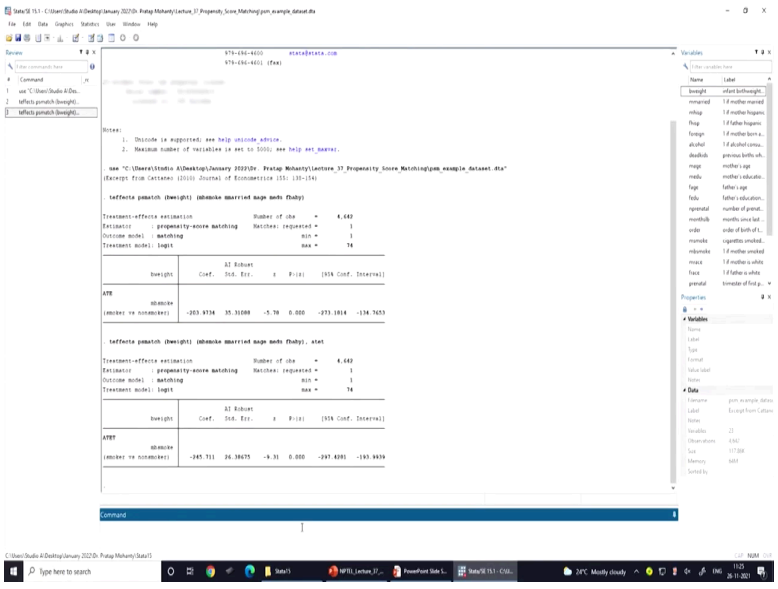
So, the same command, which I have just shown you, you can just take a note. It is in fact highlighted here. The command is t effects p effects not effect, t effects ps match propensity score match by birth weight. Birth weight is our outcome variable and by its first variable is

the treatment one then rest are being your control variables are called treatment independent variables,

So, that is important, that is the same one if you just feed it into the command window you will also get the result correctly. Now, this is on the population. We have already clarified the difference between the population as compared to the treated group. You can also find out the result with the treated group as well within the group that is called t effects know ps match,

t effects which may ps match. In this case the difference rest are the same the difference is that we need to specify which one we are going to find the result. If we are specifying within a comma followed by atet average treatment effect on the treated group, if you try to find out then we have to specify this. With this like I we can just show it over here as well. The same command we are just taking it once again and just carrying forward with comma atet,

(Refer Slide Time: 35:45)



Now, this is going to give us the result. So, within the sample where the treatment has been applied, we are comparing the same group before the treatment and after the treatment, we found that that mother who did not take did not smoke as compared to smoke those who have already those who are smoking during their pregnancy their birth weight actually reduced.

So, now, in this in the treatment group itself the impact is higher than that of the control group. We are not giving the logic behind who the difference at this moment, but it depends

on your experimentation and you will find out the result accordingly, just to compare which one is more impactful for your understanding.

And I am sure you can find out more results on this and that will be going to give you a better evaluation of large-scale data, and you will have a better publication out of it. Any sort of queries if you have done not hesitate and come back to us in writing on the query portal of NPTEL, we will be happy to deal with all those things very carefully, our team is quite active,

So, thank you very much. I expect you are participating in the next class.

Thank you, bye.