* **Research Methods Homework 3 Assignment**

Balance Table to compare Treatment and Control Groups

The treatment here is that the college is “ranked”. The control group consists of colleges that are not ranked, and we wish to compare these two groups on the three available observable characteristics.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Control (Unranked)** | **Treatment (Ranked)** | **Difference** |
| **Characteristic** | N=50 | N=50 |  |
| School Academic Quality | 0.515 | 0.466 | 0.049 |
| School Athletic Quality | 0.424 | 0.551 | -0.127\*\* |
| Proximity to Metropolitan Area | 0.360 | 0.700 | -0.340\*\*\* |

We can clearly see that the two groups significantly differ on two of the three observable characteristics – School Athletic Quality and Proximity to a Metropolitan Area. This makes direct causal inference difficult. Since the treatment group has schools with higher athletic quality and are closer to metropolitan areas, we may find the coefficient of “Ranking” to be biased as it may subsume some of these effects. For example, since treated observations are closer to metropolitan areas, and metropolitan proximity helps funding, then we might overestimate the coefficient of being ranked.

* As you know, propensity score methods are more credible when we (the researchers) can use all variables that the agents who assign treatment scan use in their assignments.

The goal of the researcher, when selecting on observables, should be to condition on all covariates that are distributed across treatment and control groups in a non-random fashion. Once we can identify all these non-random characteristics (and condition on them), we can claim that the remaining variation in the data is purely random. This implies that the units in the treatment and control group are, on average, similar on all observable characteristics except the treatment itself. Then, the effect unearthed by our analysis is solely due to the treatment, and nothing else.

When we are not able to use all the variables that the agents used to assign treatment, we are missing out on some non-random characteristic that is correlated with treatment assignment. Then, our analysis of the effect of treatment on the outcome also includes the effect of the unobservable variables. Thus, the obtained coefficient is biased.

In this setting, we do not know how the schools are ranked, and we only have access to three observable characteristics of the schools. Suppose, hypothetically, that a fourth variable in the ranking process exists – Number of past alumni playing at the big leagues. Then, it may be so that these past alumni regularly contribute and donate to the school. But since this variable is likely to be correlated with treatment assignment, we might mistakenly assume that the positive effect of the treatment on donations is due to treatment itself as we have failed to control for this fourth variable.

**Propensity Score Stacked Histogram**

**Chart, histogram

Description automatically generated**

As we can see, observations with a propensity score of greater than 0.8 are all in the treatment group. There is no overlap, i.e., these observations are always expected to be in the treatment group. Hence, I remove these observations from our analysis. The rest of the areas have some degree of overlap across both treatment and control groups.

**Blocking on the propensity score:** I create two block groups here – one with four units in each block, and the other with ten units in one block. I then use these as fixed effects in my regression next.

**Regression Results**

|  |  |  |
| --- | --- | --- |
|  | **DV:** Subsequent Year Alumni Donations | |
| Top Basketball Program Ranking | 500\*\*\* | 500\*\*\* |
|  | (.27) | (.24) |
| School Academic Quality | 102\*\*\* | 100\*\*\* |
|  | (1.9) | (.87) |
| School Athletic Quality | 47\*\*\* | 51\*\*\* |
|  | (4.1) | (1.9) |
| Proximity to Metropolitan Area | 997\*\*\* | 1,000\*\*\* |
|  | (3.4) | (1.5) |
| Constant | -.21 | .073 |
|  | (1) | (.43) |
| Fixed Effects Block at Units of: | 4 | 10 |

Standard errors in parentheses: \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

The coefficients of the covariates and the block fixed effects have very little interpretation here. This is primarily because these covariates are significantly correlated with the block fixed effects. Thus, with high collinearity, the coefficients are biased and unstable. However, we can still safely interpret the coefficient of the Basketball program ranking (treatment effect), and this is positive.

Thus, on average, keeping the covariates constant, being ranked in this basketball program meant that the school received an additional 500,000 dollars in alumni donations in the subsequent year.