Assignment 03

1. **Data set descriptions:**

The dataset comprises data on 100 colleges, with the following key characteristics:

*College Id*: A unique identifier for each college, ranging from 1 to 100.

*Academic Quality*: A score between 0 and 1 representing the academic quality of each college.

*Athletic Quality*: A score between 0 and 1 representing the quality of the college's athletic programs.

*Near Big Market*: A binary variable indicating whether a college is located near a large metropolitan area (1) or not (0).

*Ranked.2017:* A binary variable indicating whether a college was ranked in a top basketball program in 2017 (1) or not (0).

*Alumni Donations 2018*: The amount of alumni donations received in 2018.

1. **Balance Table:**

Here, the treatment is being ranked. The control is not being ranked. The covariates are

1. *Academic Quality*

2. *Athletic Quality*

3. *Near Big Market*

*Table 01: Balance Table*.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | **Control** | | **Treatment** | **Difference** |
| *Academic Quality* | 0.515 | | 0.466 | 0.049 |
|  | (0.300) | | (0.278) |  |
| *Athletic Quality* | 0.424 | | 0.551 | **-0.127\*\*** |
|  | (0.276) | | (0.281) |  |
| *Near Big Market* | 0.360 | | 0.700 | **-0.340\*\*\*** |
|  | (0.485) | | (0.463) |  |
| Observations | 50 | | 50 |  |
| Standard errors in parentheses | |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | |

Interpretations:

1. Overall, the balance table reveals notable differences between the ranked and unranked groups, especially in terms of athletic quality, proximity to big markets.
2. This indicates that these factors could be influential in a college's likelihood of being ranked in top athletic programs and in attracting higher alumni donations.
3. In sum, the balance table failed. The treatment and control groups are observably dissimilar. *The control and treatment groups are significantly different in terms of athletic quality and proximity to big markets (The differences in terms of both these are significant at 1% level*.
4. **Propensity Score Model:**

We now develop a simple logistic model of who is ranked. And then predict the propensity scores to construct a matched sample between the treatment and control group based on observables.

*Table 02: Logit Model*

|  |  |
| --- | --- |
|  | **Ranked** |
| *Academic Quality* | -0.884 |
|  | (0.780) |
| *Athletic Quality* | 1.964\*\* |
|  | (0.806) |
| *Near Big Market* | 1.615\*\*\* |
|  | (0.459) |
| Constant | -1.378\*\* |
|  | (0.645) |
| Observations | 100 |
| Pseudo-R squared | 0.141 |
| Standard errors in parentheses | |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | |

Interpretations:

1. Athletic quality and proximity to a big market are significant positive predictors of a college being ranked (Athletic Quality being significant at 5% level and proximity to big market at 1% level).
2. Academic quality inversely affects the likelihood (although not statistically significant).

**Propensity Score Predictions**

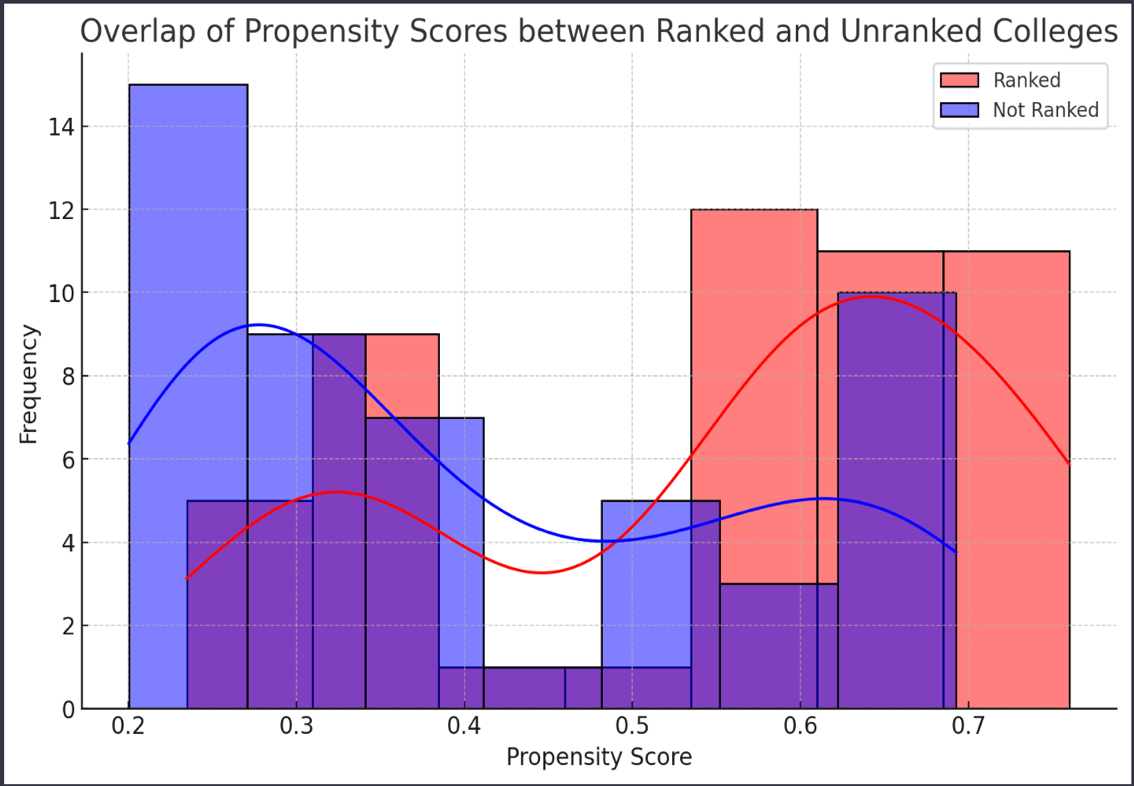
Further, we do the following to predict the propensity score:

1. The propensity scores are predicted from the above logit model.
2. This is done by predicting the probability of treatment for each observation.
3. **Stacked histograms to show overlap:**

The next step is to use stacked histograms to show the overlap in the propensity scores between ranked and unranked schools.

This will help visualize the distribution and determine if there are distinct differences between the two groups.

*Figure 01: Stacked Histogram (Treatment and Control)*

****

Observations from the Histogram:

1. *Overlap Region*: There is a noticeable overlap in the propensity scores between the two groups. This suggests that there are colleges in both the ranked and unranked categories that have similar probabilities of being ranked based on the model.
2. *Distinct Groups*: Despite the overlap, certain portions of the histogram show distinctions between ranked and unranked colleges. Ranked colleges tend to have higher propensity scores, as indicated by the red histogram leaning towards the right.
3. B**locking on the propensity score:**

* The dataset has been divided into blocks based on the propensity scores.
* Each block contains a group of colleges with similar likelihoods of being ranked, as determined by the logistic regression model.
* This blocking method is a key step in propensity score matching, allowing for a more controlled comparison between ranked and unranked colleges.

1. **Treatment effect of being ranked on alumni donations:**

The next and final step in the assignment is to analyze the treatment effect of being ranked on alumni donations. This involves:

1. Conducting a statistical analysis considering block-fixed effects and other covariates.
2. Determining if there's a significant difference in alumni donations between ranked and unranked colleges within each block.

*Table 03: Regression table for Alumni Donations*

|  |  |
| --- | --- |
| **OLS** | **Donations** |
|  |  |
| Ranked dummy | 500.5\*\*\* |
|  | (0.264) |
| Academic Quality | 101.4\*\*\* |
|  | (1.726) |
| Athletic Quality | 47.19\*\*\* |
|  | (3.636) |
| Near Big Market | 997.3\*\*\* |
|  | (2.994) |
| Block Fixed Effects (Propensity Score Blocks) | 25 |
| Constant | -0.137 |
|  | (0.933) |
| Observations | 100 |
| R-squared | 1.000 |
| Standard errors in parentheses |  |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 |  |

Interpretations:

1. *Academic Quality*: Significant (p < 0.001) with a coefficient of 101.4. This suggests that higher athletic quality is associated with increased alumni donations.
2. *Athletic Quality*: Highly significant (p < 0.001) with a coefficient of 47.19 This suggests that higher athletic quality is associated with increased alumni donations.
3. *Near Big Market*: Highly significant (p < 0.001) with a coefficient of 997.3, indicating a strong positive association between being near a big market and higher alumni donations.
4. *Ranked 2017*: Extremely significant (p < 0.001) with a coefficient of 500.5, indicating a very strong positive association between being ranked and increased alumni donations.
5. R-squared: 1.000, suggesting that the model perfectly explains the variance in alumni donations. However, such a high R-squared might indicate overfitting.
6. The analysis strongly suggests that being ranked in a top program has a substantial positive effect on alumni donations, even after controlling for academic and athletic quality, proximity to big markets, and block-fixed effects.
7. Blocks: Suggests that the blocking based on propensity score does not significantly impact the variance in alumni donations in this model.