**Answers to Study Questions for Chapter 4**

1. Table 4.1 shows the (a) selection bias and (b) standardized bias for each of the 10 covariates before any propensity score adjustment. Notice that all of the covariates have a standardized bias that is greater than 0.10, suggesting that they are not balanced.

**Table 4.1** Selection bias and standardized bias before propensity score adjustment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Covariate | Treatment Mean | Control Mean | Selection Bias | Standardized Bias |
| Propensity score | 0.583 | 0.179 | 0.405 | 1.564 |
| HSGPA | 2.919 | 3.388 | -0.469 | -0.887 |
| ACT | 20.100 | 23.100 | -3.000 | -0.911 |
| Extraversion | 2.580 | 2.437 | 0.143 | 0.180 |
| Agreeableness | 3.147 | 2.800 | 0.347 | 0.764 |
| Neuroticism | 1.743 | 1.981 | -0.238 | -0.373 |
| IMaccomplish | 3.508 | 2.914 | 0.594 | 0.426 |
| IMexperience | 2.742 | 2.146 | 0.595 | 0.457 |
| Loneliness | 0.758 | 1.087 | -0.329 | -0.646 |
| Caucasian | 0.600 | 0.786 | -0.186 | -0.373 |
| AfricanAm | 0.267 | 0.143 | 0.124 | 0.275 |

1. After computing propensity scores from all 10 covariates in the *First Year Seminar* dataset, treatment cases (*Univ101* = 1) were matched to control cases (*Univ101* = 0) using paired, nearest neighbor matching, which ordered propensity scores from highest to lowest, without replacement or a caliper. Table 4.2 shows the (a) selection bias, (b) standardized bias, and (c) percent bias reduction for each covariate after propensity score matching.

**Table 4.2** Selection bias, standardized bias, and percent bias reduction after propensity score matching.

|  |  |  |  |
| --- | --- | --- | --- |
| Covariate | Selection Bias | Standardized Bias | Percent Bias Reduction |
| Propensity score | 0.224 | 86.568 | 44.637 |
| HSGPA | -0.199 | -37.553 | 57.641 |
| ACT | -0.500 | -15.178 | 83.333 |
| Extraversion | 0.073 | 9.240 | 48.667 |
| Agreeableness | 0.170 | 37.453 | 50.962 |
| Neuroticism | -0.027 | -4.182 | 88.800 |
| IMaccomplish | 0.317 | 22.725 | 46.693 |
| IMexperience | 0.358 | 27.515 | 39.800 |
| Loneliness | -0.087 | -17.025 | 73.642 |
| Caucasian | -0.033 | -6.690 | 82.051 |
| AfricanAm | 0.000 | 0.000 | 100.000 |

* 1. Only *Extraversion*, *Neuroticism*, *Caucasian*, and *African American* are balanced after propensity score matching. All of the other covariates are still unbalanced (i.e., *SB* > 0.1) This suggests that either the propensity score model or the matching method was not sufficient. We may improve these results by using other types of matching methods (e.g., caliper matching, optimal matching, etc.).
	2. Propensity score matching reduced the selection bias for *ACT*, *Neuroticism* *Caucasian*, and *African American* by 80% or more. However, the percent bias reduction for the other six covariates was less than 80%.
1. Propensity score matches may vary across iterations (even when using the same propensity score model and matching method) when more than one treatment or control case has the same propensity score. Therefore, the results provided below may be slightly different from the results you obtain. However, the models and methods should be robust enough to make the same inferential conclusions. After matching on propensity scores:
2. There is no significant difference in first-year grade point average between those who did (*M* = 2.632, *SD* = 0.760) and did not (*M* = 2.819, *SD* = 0.810) participate in the first year seminar program, *t*(58) = -0.922, *d* = -0.238, *p* = 0.360.
3. There is no significant difference in the percent of students who returned to college for a second year between those who did (83% of the people in the treatment group remained in college) and did not (77% of those in the control group remained in college) participate in the first year seminar program, *χ*(1) = 0.417, *OR* = 1.522, *p* = 0.519.
4. Before matching on propensity scores:
5. Using the original data before matching, students who participated in the first year seminar program had significantly lower first-year grade point averages (*M* = 2.632, *SD* = 0.760) than those who did not (*M* = 2.999, *SD* = 0.747), *t*(98) = -2.237, *d* = -0.488, *p* = .028. Before matching on propensity scores the selection bias makes the treatment appear to have a negative effect. However, after matching on propensity scores, there is no treatment effect.
6. Using the original data before matching, there is no significant difference in the percent of students who returned to college for a second year between those who did (83% of the people in the treatment group remained in college) and did not (76% of those in the control group remained in college) participate in the first year seminar program, χ(1) = 0.971, *OR* = 1.731, *p* = .325. There is not a significant difference between the two groups on first-year retention before or after matching on propensity scores. However, the effect size decreases slightly (from 1.731 to 1.522) after matching on propensity scores.
7. Using the results from Problem 3, Rosenbaum’s Sensitivity tests for Wilcoxon Signed Rank and Hodges-Lehmann were run for first year grade point average. Table 4.3 shows the upper and lower bounds for each **.
8. If the unbiased estimate is significant and the *p* value becomes non-significant within the bounds (e.g., upper bound > .05), the result from the treatment estimate is sensitive when ** < 2.0. If the unbiased estimate is non-significant, as in this example (i.e., ** = 1, *p* = 0.56), but the *p* value within the bounds becomes significant (i.e., bounds include .05) when ** = 1.1, the result indicates that the treatment estimate using the propensity score model is sensitive to hidden bias.
9. In this example, the treatment effect estimation becomes sensitive to bias with only a 0.1 increase in gamma (i.e., ** ≥ 1.1). Therefore, the results from problem 3 may not be valid if there are any other confounding variables contributing to selection bias that were not included in our propensity score model. In addition, we can also use the Hodges-Lehmman point estimates (See Table 4.3) for the sign rank test with the R function *hlsens* for additive effect due to treatment. As show in Table 4.3, the difference in medians between treatment and control groups in first year GPA without hidden bias is -0.31007. When ** = 1.9, the bounds (-0.71007 to 0.08993) include zero, which indicates that the treatment effect estimate is still sensitive to hidden bias with ** < 2.0.

**Table 4.3** Upper and lower bounds for Rosenbaum’s Sensitivity tests for first year GPA.

|  |  |  |
| --- | --- | --- |
|  | Wilcoxon Signed Rank P-Value | Hodges-Lehmann Point Estimate |
| Gamma | Lower Bound | Upper Bound | Lower Bound | Upper Bound |
| 1.0 | 0.0560 | 0.0560 | -0.31007 | -0.31007 |
| 1.1 | 0.0347 | 0.0858 | -0.41007 | -0.21007 |
| 1.2 | 0.0214 | 0.1217 | -0.41007 | -0.21007 |
| 1.3 | 0.0131 | 0.1625 | -0.51007 | -0.11007 |
| 1.4 | 0.0080 | 0.2070 | -0.51007 | -0.11007 |
| 1.5 | 0.0049 | 0.2539 | -0.51007 | -0.11007 |
| 1.6 | 0.0030 | 0.3022 | -0.61007 | -0.01007 |
| 1.7 | 0.0018 | 0.3508 | -0.61007 | -0.01007 |
| 1.8 | 0.0011 | 0.3989 | -0.61007 | -0.01007 |
| 1.9 | 0.0007 | 0.4457 | -0.71007 | 0.08993 |
| 2.0 | 0.0004 | 0.4909 | -0.71007 | 0.08993 |