**Answers to Study Questions for Chapter 2**

1. Covariates should be (a) statistically related to group assignment (i.e., treatment vs. control groups), (b) statistically related to the outcome, and (c) theoretically influential in how a participant selects or is assigned to a treatment group.
2. (a) High school GPA (*HSGPA*), ACT, Agreeableness, Internal Motivation to Accomplish Things (*IMaccomplish*), Internal Motivation to Experience Stimulation (*IMexperience*), Loneliness, and Caucasian have the strongest relationships with the treatment condition (*Univ101*). See Table 2.1.

(b) High school GPA (*HSGPA*), ACT, Agreeableness, Caucasian, and African American have the strongest relationships with *FirstYrGPA*. None of the covariates has a strong relationship with *EnrollYr2.* See Table 2.1.

**Table 2.1** Statistical relationships between each covariate and treatment and outcomes.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Univ101* | *FirstYrGPA* | *EnrollYr2* |
| *HSGPA* | *t* = -2.845, *p* = .005 | *r* = .481, *p* < .001 | *t* = 0.271, *p* = .787 |
| *ACT* | *t* = -3.522, *p* = .001 | *r* = .261, *p* = .009 | *t* = -0.904, *p* = .368 |
| *Extraversion* | *t* = 1.090, *p* = .279 | *r* = -.033, *p* = .747 | *t* = 0.857, *p* = .394 |
| *Agreeableness* | *t* = 2.795, *p* = .006 | *r* = .189, *p* = .059 | *t* = 1.721, *p* = .088 |
| *Neuroticism* | *t* = -1.381, *p* = .171 | *r* = .103, *p* = .309 | *t* = -.725, *p* = .470 |
| *IMaccomplish* | *t* = 2.289, *p* = .024 | *r* = .174, *p* = .083 | *t* = 1.043, *p* = .299 |
| *IMexperience* | *t* = 2.252, *p* = .027 | *r* = .044, *p* = .664 | *t* = 1.154, *p* = .251 |
| *Loneliness* | *t* = -2.812, *p* = .006 | *r* = .007, *p* = .948 | *t* = -1.279, *p* = .204 |
| *Caucasian* | *χ2* = 3.675, *p* = .055 | *χ2W* = 8.442, *p* = .004 | *χ2* = 0.001, *p* = .974 |
| *AfricanAm* | *χ2* = 2.181, *p* = .140 | *χ2 W* = 6.189, *p* = .013 | *χ2* = .001, *p* = .980 |

Note: Independent samples *t*-tests were used to examine relationships between the continuous covariates and the treatment condition (*Univ101*) and categorical outcome variable (*EnrollYr2*). Chi Square tests were used to examine relationships between the categorical covariates and the treatment condition (*Univ101*) and categorical outcome variable (*EnrollYr2*). Pearson correlations were used examine relationships between the continuous covariates and the continuous outcome variable (*FirstYrGPA*). Wald Chi Squares (from logistic regressions) were used examine relationships between the categorical covariates and the continuous outcome variable (*FirstYrGPA*).

1. College major, disability status, whether or not the student is a first-generation college student, whether or not the student is receiving financial aid, and conscientiousness may be other good covariates.
2. Propensity scores for the first 10 cases for each of the four models are in Table 2.2.
   1. Model A includes only the covariates that are significantly related to the treat­ment condition: High school GPA (*HSGPA*), ACT, Agreeableness, Internal Motivation to Accomplish Things (*IMaccomplish*), Internal Motivation to Experience Stimulation (*IMexperience*), and Loneliness.
   2. Model B includes only the covariates that are significantly related to college performance (*CollGPA*): High school GPA, ACT, Caucasian, and African American.
   3. Model C includes only the covariates that are significantly related to both the treatment condition and college performance: High school GPA and ACT
   4. Model D includes the covariates that are significantly related to either the treatment con­dition or college performance: High school GPA, ACT, Agreeableness, Internal Motivation to Accomplish Things, Internal Motivation to Experience Stimulation, Loneliness, Caucasian, and African American.

**Table 2.2** Propensity scores from propensity score Models A, B, C, and D.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Model A | Model B | Model C | Model D |
| 1 | .041 | .093 | .095 | .038 |
| 2 | .041 | .112 | .117 | .038 |
| 3 | .028 | .080 | .085 | .024 |
| 4 | .070 | .081 | .085 | .064 |
| 5 | .029 | .068 | .072 | .024 |
| 6 | .661 | .601 | .598 | .691 |
| 7 | .249 | .435 | .427 | .279 |
| 8 | .421 | .494 | .525 | .351 |
| 9 | .159 | .129 | .133 | .157 |
| 10 | .109 | .102 | .106 | .100 |

1. Using the same covariates that were used in Model C from the previous problem, the propensity scores for the first 10 cases using the
   1. classification tree (CART) method and
   2. generalized boosted modeling (GBM) are listed in Table 2.3.
   3. Notice that the propensity scores estimated from each algorithm are slightly different. Propensity scores from the CART have the least variation (i.e., those for cases 1, 2, 3, 4, 5, 9, and 10 are all .109), while propensity scores from the GBM procedure have the most variability (i.e., each score is unique). The propensity scores estimated from the logistic regression are most similar to those from the CART, and the propensity scores from the CART are least similar to those from the GBM.

**Table 2.3** Propensity scores estimated from CART and GBM algorithms in *R*.

|  |  |  |
| --- | --- | --- |
| Id | CART | GBM |
| 1 | .109 | .308 |
| 2 | .109 | .087 |
| 3 | .109 | .062 |
| 4 | .109 | .056 |
| 5 | .109 | .053 |
| 6 | .625 | .502 |
| 7 | .571 | .349 |
| 8 | .571 | .425 |
| 9 | .109 | .090 |
| 10 | .109 | .073 |

Note: Covariates include High school GPA and ACT.