

Does participation in the Smith Summer Research Fellows (SURF) program increase your odds of attaining an advanced degree?

Tanya Hakim (2012), Nicholas Horton, Cate Rowen, and Minh Ly

Department of Mathematics and Statistics & Office of Institutional Research and Educational Assessment, Smith College



Introduction:

The Smith Summer Research Fellows (SURF) program was started in 1967 and creates opportunities for students to participate in collaborative research with faculty. One of the major goals of the program is to provide the necessary research experience and support for women to succeed in pursuing a post-graduate degree, particularly in the sciences. The program requires financial support from the college to provide stipends for the participants, and in turn, it is crucial to understand whether the program increases the likelihood participants will pursue an advanced degree in the sciences. In this study, we analyze data from a survey of Smith College graduates, some of whom have participated in the program and the rest who have not, in order to see whether participation in the program is correlated with increased probability of pursuing advanced degrees, especially in the sciences.

Methods:

About the Dataset

- Collected data comes from the Smith College Alumnae Survey and BannerWeb. More details can be found in [1] and [2].

Predictor Variables

- SRF Status:** Indicates whether the student participated in SURF or not
- Race/ethnicity:** Describes race/ethnicity background of the student. Categories are Underrepresented Minority, White, International and Asian
- Reader Rating:** Describes admission's office rating of the student's application (ratings of 1 are best, and 8 worst). This data was missing for 136 subjects
- Total SAT Score:** Describes the total SAT score for the student. Includes verbal score and math score with both being scored out of 800 for a potential total of 1600 for the combined score. For 8 students who did not have a total SAT score but had a composite ACT score, their scores were imputed from the composite ACT to the total SAT using a concordance table available through College Board [3]. This data was missing for 21 subjects.
- Graduation Year:** Describes the year of graduation for the subject

Outcome Variable

- Degree type (type*):** This dichotomous variable describes the highest degree the subject attained. The two possible groups for the variable were attainment of a Bachelor's or below OR any advanced degree (Master's and above)

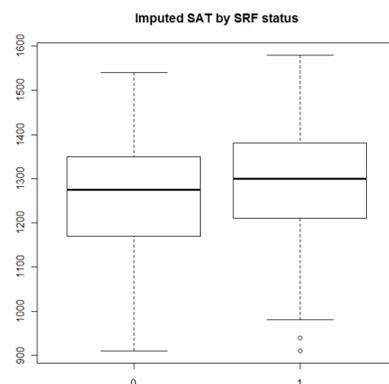
Statistical Methods

Missing Data

- Gradyr is a statistically significant predictor for missingness of SAT score ($p=0.013$)
- Mean for those missing total SAT score was 2003, while the mean for those not missing the total SAT score was 2001. This suggests that the data for total SAT is missing for more recent years
- Those missing values for total SAT scores tended to have better reader ratings, though this was not statistically significant ($p=.068$)
- Gradyr is also a statistically significant predictor for missingness of rrating ($p<0.001$). In the years 1999 and below, only 5 rratings were reported, while in the years 2000 and above, only 4 rratings were missing
- To best utilize the available information, we fit three models: one with just SAT, one with just rrating and another using all predictors. To simplify interpreting these models, those students missing both total SAT and rrating were dropped

Multiple Logistic Regression Model

- Outcome is either a graduate degree or bachelors and below
- Primary predictor variable is whether the student participated in SURF
- Controlling for confounding variables SAT score, evaluative rating of the student by the admissions office, ethnicity and graduation year



The boxplot to the left shows total SAT scores for who participated in SURF, indicated by 1, and those who did not, indicated by 0. SURF participants tended to have higher SAT scores than non-participants, and in turn, it was necessary to control for total SAT score. Our results were still statistically significant even when controlling for SAT score.

Propensity Score Models

- Each subject given a propensity score describing estimated probability the subject would be a participant in the SURF program on the basis of his or her own individual variables (i.e. ethnicity, total SAT score, rrating)
- Participants of the SURF program are matched with students not participating in the SURF program who have similar propensity scores
- Matching allows for minimization of differences that may exist between participants of the program and non-participants [4]
- Can control for the propensity score in a multiple logistic regression model
- Can also match based on the propensity score. Analysis can be performed on the matched samples in order to see whether a correlation exists between SURF participation and increased likelihood of successfully pursuing an advanced degree

Results:

Descriptive Statistics

- SRF Status:** 36.1% of subjects ($n=149$) did not participate in SURF, while 63.9% ($n=264$) did participate
- Race/ethnicity:**
 - 7.99% of subjects ($n=33$) Asian,
 - 6.54% ($n=27$) International,
 - 7.02 ($n=29$) Underrepresented Minorities
 - 78.5% ($n=324$) White
- Reader Rating:** The five number summary is (1, 3, 4, 5, 8)
- Total SAT Score:** The five number summary is (910, 1200, 1290, 1370, 1580). Relatively normal distribution though skewed slightly left.
- Graduation Year:** The five number summary is (1995, 1998, 2001, 2004, 2007). The distribution of graduation years is relatively uniform

Multiple Logistic Regression Model and Multiple Logistic Regression Model Controlling for Propensity Score

| | SAT only Mean (95% Confidence Interval) | Reader Rating Only Mean (95% CI) | Both SAT and Reader Rating Mean (95% CI) |
|---|---|-------------------------------------|--|
| Participated in SURF (Multiple Logistic Regression) | 1.9 (1.2-3.0) $n=384$ | 2.0 (1.2-3.5) $n=269$ | 2.0 (1.1-3.5) $n=247$ |
| Participated in SURF (Controlling for Propensity Score) | 1.9 (1.2-2.9) $n=384$ | 2.1 (1.4-3.2) $n=269$ | 1.9 (1.2-3.0) $n=247$ |

Matched Analyses

- SURF participants were more likely to complete an advanced degree relative to non-SURF participants ($p=0.059$, $t=-1.889$)
- SURF status is not as statistically significant in the matched analyses using propensity scores compared to the results for the multiple logistic regression model
- This difference likely due to the students dropped when matches were made. Students in the control group will be dropped from the analysis who do not have a similar match in the treatment group and vice versa

Discussion:

- Participation in SURF appeared to be associated with greater completion of graduate degrees
- Important to remember that graduate school not necessarily the only indicator of success in a field and in one's career.
- In future, our outcome variable could be coded to separate those pursuing advanced degrees in the sciences from the other disciplines to isolate whether the program specifically increases the likelihood SURF participants will pursue graduate degrees in the sciences in particular

References

- Lopatto, David and Carol Trosset. "Report on the Smith College Alumnae Survey." Clark Science Center Director's Office, Smith College. (2008).: 1-26. Web. 1 April 2010 <<http://www.science.smith.edu/departments/director/extranet/documents/ReportontheSmithCollegeAlumnaeSurveyDL2.pdf>>.
- BannerWeb. Office of the Registrar, Smith College.
- ACT – SAT Concordance Tables. College Board. Web. 1 April 2010. <<http://professionals.collegeboard.com/data-reports-research/sat/sat-act>>
- Reiter, Jerome. "Using Statistics to Determine Causal Relationships." *American Mathematical Monthly* 107.1 (January 2000): 24-32. Print.