The Impact of Supplemental Instruction on Student Success

Aguilar R., Donatello R. Dr.PH
California State University, Chico

Introduction

Universities want to reduce the time it takes for students to graduate and to minimize the dropout rate. To do this, universities invest in methods to reduce the rate at students fail classes. One of these methods is to provide students with Supplemental Instruction (SI) for classes with high failure rates. SI is a voluntary, confidential, and peer-led academic support model.

Prior research at the individual course level (i.e. just CHEM 111) has shown that students who attend SI pass these classes at a higher rate than those who do not attend SI. However, these analyses tend to stop after only comparing certain groups of interest such as First gen vs non-first gen.

For observational studies like these, bivariate analyses should be considered a starting point and mostly for exploratory analyses only. Confirmatory analyses should be multivariable models that control for other characteristics that are known to be associated with the outcome.

This research provides an institutional level analysis of the effectiveness of SI on the likelihood a student will pass a class after controlling for other factors known to be related to student success across all classes and all years. Several multivariable statistical models including logistic regression, random forest, stochastic gradient boosting were used to model the likelihood a student will pass a class. Models were compared on their accuracy and predictive ability.

Methods

• Deidentified data was provided by CSU, Chico’s Office of Institutional Research.
• Univariate and bivariate exploratory analysis was conducted to explore possible relationships between variables and identify any irregularities in the data.
• Several multivariable and machine learning models were fit to predict a DWF grade
  – Logistic regression
  – Random forest
  – Stochastic gradient boosting
• The synthetic minority over-sampling technique (SMOTE) was used for the random forest and stochastic gradient boosting models due to class imbalances.
• To avoid overfitting, cross-validation techniques were used with a 70% training sample and 30% testing.
• Propensity score matching was conducted to create a matched subsample of students who have similar likelihood to attend SI.
• This method is used to account for any baseline differences in likelihood of attending SI.

Sample Characteristics

• SI was offered in 132 classes over 5 years (Fall 12 – Fall 17)
• 12,730 observations on 8,164 regular semester, undergraduate students attending classes offering SI
• SI was available for 17 courses across departments including:
  – Biological Sciences
  – Chemistry & Biochemistry
  – Communication Arts & Sciences
  – Computer Science
  – Finance and Marketing
  – History
  – Mechanical & Mechatronic Engineering
  – Sustainable Manufacturing
  – Multicultural & Gender Studies
  – Political Science

Predictive Modeling

Variable importance is a way of comparing the predictive ability of covariates across different models. The higher the importance, the greater the association between the covariate and outcome

Model performance

Models were compared using accuracy (overall correct prediction), sensitivity (correctly predicting a DWF), specificity (correctly predicting an ABC) of the 30% testing sample.

3. Logistic Regression
   - Model: 0.79
   - Sensitivity: 0.81
   - Specificity: 0.77

4. Random Forest
   - Model: 0.79
   - Sensitivity: 0.78
   - Specificity: 0.74

5. Gradient Boosting
   - Model: 0.80
   - Sensitivity: 0.84
   - Specificity: 0.77

GBM was the best performing model on these metrics, but all models perform similarly.

Results

Multivariable Logistic Regression Modeling

• OR: Odd Ratios (OR) indicate the multiplicative effect that characteristic has on the probability of a DWF grade while holding all other characteristics constant
• OR < 1 indicates a lower chance of DWF, >1 indicates a higher chance

- Students who go to SI have almost half the odds (0.51) of getting a DWF grade compared to students who did not attend SI.

After controlling for SI, historically underserved students (students of color, Pell eligible) are equally likely to pass a class compared to their counterparts.

Improvements are seen across all colleges except HFA.

Marginal Effects (Model Predicted Probabilities of DWF)

As academic preparedness (HSEI) and overall cumulative GPA increase, the predicted probability of receiving a DWF grade decreases, but at different rates for those that attend SI and those that do not.

Matched Sample Comparison

• T-tests for a difference in DWF rate by SI attendance was done on the matched sample, and on the two independent groups.
• Results are similar, on average the SI group passed their classes at a higher rate than those who did not attend SI.
• Independent Groups: -0.11 (95% CI: -0.13, -0.09)
• However, the examination of diagnostic plots showed that the treatment and control groups were not matched well.
• More research is needed to provide evidence of a causal relationship.

Implications

This work contributes to the growing body of literature indicating Supplemental Instruction is an effective student success program.

We hope this research aids our university administration in future funding decisions regarding student success programs.

Acknowledgements

Dr. Erik Weisanger (Chemistry) for collecting and providing the data, for the interesting research question, for providing subject matter consulting, and for overall being a champion of Supplemental Instruction at Chico State.

References