

DRAFT Proposal for Developing a More Robust Enrollment Analysis and Forecasting Model

October 13, 2020

Purpose and Problem Statement. The Office of Institutional Research and Assessment proposes to partner with students and faculty in the Statistics and Operations Research program to integrate data science techniques into our ongoing analyses and projections of enrollments. To a large extent, our revenues are generated by student credit hour enrollments. In a large, decentralized institution with many programs, it is difficult to project future semester enrollments and therefore the funding that we might generate. Our current projections use simple techniques such as ratio analyses in spreadsheets and linear regression. Our goal is to create a predictive model built upon an algorithm that takes into account variations in student enrollments and coursetaking patterns in a more sophisticated way and results in a more accurate prediction of credit hour production and revenue.

In addition to predicting revenue from enrollment change, a new and enhanced credit hour projection model will allow us to examine other more complex questions. Examples include how shifts in patterns of student choices of majors or the requirements of the new Gen Ed curriculum might impact credit hour enrollment change and funding; and how the UNC System's new policy requiring campuses to award course credit for AP scores of 3 or higher could change enrollments in introductory classes and subsequent courses in that discipline. There are many other questions we could explore that would benefit from a more advanced approach to analysis.

We envision forming an analytics team of STOR graduate students and faculty that would include Chris Eilers and Lynn Williford, thus taking advantage of the advanced data science expertise of the STOR students and faculty and the subject matter expertise and data knowledge of Institutional Research and Assessment. The Office of Institutional Research and Assessment is responsible for the protection of the data and privacy of students and work to ensure that the data used in the project are de-identified. The STOR graduate students and faculty partners will sign a non-disclosure statement (which all OIRA employees sign before working with data), and the data will be kept securely on a password protected site managed by OIRA or other secure arrangement. It is important that the work products cannot be used for research or shared beyond the group.

Data Available for Analysis

We propose to limit this initial study to undergraduate students. Below is a rough list of data that are available for analysis. This list may change as the study progresses. The unit of analysis is course credit hours taught during fall and spring terms in academic years 2015 to 2019. The data source is the Census files for which OIRA serves as the campus data steward. One row per student course enrollment, with the following variables:

- Credit hour and course characteristics:
 - Number of credit hours attempted and earned for that enrollment
 - Course subject
 - Course number and section number
 - Department/school course owner
 - Course level of instruction (on the 12 cell model)
 - Cost category of credit hours (on the 12 cell model)
 - Fundability of credit hours generated based on the course status (reasons for non-fundability include: offered by an off-model program, course funded externally, etc.)
 - Term taught
 - Mode of instruction (face-to-face vs. remote)
- Characteristics of enrollee

- Fundability of credit hours generated based on the student's status – Y/N
- Academic level: undergrad, master's, doctoral, professional
- Degree or non-degree
- Year in school
- Entered as transfer or first year
- Rural*
- Needy*
- First generation*
- Residency
- Major
- Cum GPA
- Test scores and credits earned upon entry
- Total credit hours attempted and earned that term
- Tuition charged that term
- Term of first enrollment
- Terms in Residence
- Returned year 2, year 3, year 4, year 5, year 6
- Graduation in 2-3 years (transfers), 4 years, 5 years, 6 years

*The file will be de-identified by removing name, PID, and any other direct identifiers. The demographic data elements (rural, needy, first gen, all (Y/N) should be combined to create a single scale representing socio-economic status that can be used in the analysis in place of the individual variables. This study will use this scale as one of many predictors of credit hour enrollment outcomes and funding. We will not use it to examine group differences in performance that might result in stereotyping.