Predicting Institutional Retention and Graduation Rates Using IPEDS and WSCUC GRD Data

Problem Statement

• Retention and graduation rates are key performance indicators (KPIs) used in college rankings, accreditation and assessment reports, and national accountability systems.

• Retention and graduation rates are used increasingly as part of state-support funding formulas for public colleges and universities.
Background

• Today’s study builds on a research project completed at the AIR/NCES National Data Institute in Washington D.C. (2013), and subsequent research by Pike & Graunke (2015).

• Objective of NDI work was to produce two institution-level prediction models (for retention and graduation rates) using IPEDS data and standard regression methods.

Purpose of Study

• Develop prediction models for institutional retention and graduation rates.
• Develop models for public and private institutions.
• Utilize federally reported higher education data (IPEDS) that is readily accessible to higher education analysts.
• Offer online prediction tool to perform “what-if” scenarios associated with changes in student and institutional attributes in order to inform enrollment management and strategic planning for a wide-range of institutions.
### Relevant Previous Literature


### Data Sources and Data Elements

- **Sources**
  - IPEDS
  - WSCUC GRD (future research)

- **Institution cohorts**
  - New full-time first-year students
  - Retention cohort: academic year 2012-13
  - Graduation cohort: academic year 2008-09

- **Institutional data elements (predictors)**
  - Demographics
  - Selectivity
  - Enrollment
  - Financial aid
  - Finances
  - Student-faculty ratio
Data Sources and Data Elements

- Demographics
  - Student ethnicity (% composition), average student age
- Admission selectivity/Student success
  - Test score (%tiles), historical 6-yr graduation rate, freshmen retention
- Enrollment
  - Freshmen class size, out of state %, part-time %, rural %, distance education course enrollment
- Financial aid
  - Federal grant support (%), net price of attendance (high-income students), Pell students (%)
- Finances
  - Instructional expenses (% of total, per FTE)
- Student-faculty ratio
  - student-faculty ratio

Institution Selection in Data Center

<table>
<thead>
<tr>
<th>EZ Group Selection</th>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total IPEDS First Look Universe</td>
<td>7,252</td>
<td></td>
</tr>
<tr>
<td>US, Public/Private not-for-profit 4-year above</td>
<td>1,814</td>
<td></td>
</tr>
<tr>
<td>Has full-time undergraduates,</td>
<td>1,725</td>
<td></td>
</tr>
<tr>
<td>No completely online campuses</td>
<td>1,719</td>
<td></td>
</tr>
<tr>
<td>Final selection</td>
<td>1,719</td>
<td></td>
</tr>
</tbody>
</table>
Variable Selection Framework

Institutional Characteristics
- Demographics
- Selectivity
- Enrollment
- Financial Aid
- Finances
- Student-faculty Ratio

Retained Rates

Graduation Rates

Resulted in retrieval of >250 IPEDS variables

Data Management Tasks

• Exploratory data analysis
  – Variable selection (bivariate correlation matrix)
  – Variable coding (continuous vs. dummy/binary)
  – Missing data imputation
  – Re-scaling of variable(s) to ease coefficient interpretation
    • E.g., First-Time Freshmen Headcount per 100, Instructional Expenses per 1,000-$

• Regression models
  – Preliminary model fit (linear/logit estimation, block entry, $R^2$)
  – Collinearity assessment
  – Identification of statistical outliers (residuals, centroid distance)
Data Management Tasks

- Imputation Example: SAT scores for cases with missing data
  - Regress SAT 25th Total Score on % New FR Asians, % under age 18, % over age 24, recent 6-yr grad rate, % Pell, FR retention rate, % admitted, enrollment yield.
  - Use regression coefficients to impute SAT scores for missing cases

**Model Summary**

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.866</td>
<td>.742</td>
<td>.861</td>
<td>3.01421</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Enrollment, Admitted/Pct, USAgeL18, Pell/Pct, NewFRAsians/Pct, USAgeGE24, Full-time retention rate 2013 (EfD013D_R1), Grad6Yr

**Unit Redemption Rate**

- **Data source**
  - WASC member institutions

- **Date elements**
  - Academic year completed credit hours of enrolled UG students (U)
  - Completed *institutional* credit hours of degree recipients for AY (R)

- **Unit redemption rate (URR):** \[ \frac{R}{U} \]

- **Pipeline effect**
  - Annual enrollment drop/rise increases/decreases URR
  - Smoothed effect if based on multiple-year U’s
‘Absolute’ Graduation Rate

- Data source
  - WASC member institutions

- Date elements
  - Inst. credits taken by dropouts, number of dropouts (past 5 years)
  - Completed institutional credit hours of degree recipients for AY, number of graduates for AY
  - Average credits per dropout (d), average credits per graduate (g)

- Unit redemption rate (URR): \( \frac{R}{U} \)

- Absolute graduation rate (AGR): \( \frac{R}{R + \frac{(U - R)}{d/g}} \)

The URR and AGR Rates

- Unit redemption rate (URR)
  - Approximation of proportion of institutional credits completed (all UG students) that ‘lead’ to a UG degree (all UG degree recipients) for AY

- Absolute graduation rate
  - Adjusts IPEDS-reported rate by average credits of dropouts to average credits of degree recipients (weighted by time)
IPEDS versus URR and AGR Rates

• Predicting URR and AGR rates
  – Model estimation with IPEDS data
  – Model estimation with IPEDS and WASC-requested member institution data

• Identifying key factors associated with graduation-rate metrics
  – Compare factor set for IPEDS rate versus AGR

Questions & Comments

Presentation at
http://www.unr.edu/ia/research

Contact:
serge@unr.edu
jstanley@hawaii.edu