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

Serge Herzog, PhD
Director, Institutional Analysis
Consultant, CRDA StatLab
University of Nevada, Reno
serge@unr.edu

John Stanley, MEd
Associate Director
Office of Institutional Effectiveness
University of Hawaii – West Oahu
jstanley@hawaii.edu

CAIR Conference, November 16-18, 2016, Los Angeles, CA
Introduction

- 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

**EZ-Group Selection**

<table>
<thead>
<tr>
<th>1) Total IPEDS First Look Universe,</th>
<th>7,252</th>
<th>3) Has full-time undergraduates,</th>
<th>1,725</th>
</tr>
</thead>
<tbody>
<tr>
<td>2) US, Public/Private not-for-profit four-year above,</td>
<td>1,814</td>
<td>4) No completely online campuses</td>
<td>1,719</td>
</tr>
</tbody>
</table>

Final release data used; retrieved 10/26/2016
Variable Selection Framework

Institutional Characteristics

- Demographics
- Selectivity/Student Success
- Enrollment
- Financial Aid
- Finances
- Student-faculty Ratio

Retention Rates (2012 cohort)
Graduation Rates (2008 cohort)

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

<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>.896a</td>
<td>.802</td>
<td>.801</td>
<td>3.01821</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>41.589</td>
<td>1.241</td>
<td>33.506</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>NwFRAsianPct</td>
<td>.115</td>
<td>.015</td>
<td>.113</td>
<td>7.876</td>
</tr>
<tr>
<td></td>
<td>UGAgeLt18</td>
<td>.037</td>
<td>.017</td>
<td>.026</td>
<td>2.190</td>
</tr>
<tr>
<td></td>
<td>UGAgeGt24</td>
<td>-.018</td>
<td>.007</td>
<td>-.038</td>
<td>-2.503</td>
</tr>
<tr>
<td></td>
<td>Grad6Yr</td>
<td>.125</td>
<td>.010</td>
<td>.331</td>
<td>12.892</td>
</tr>
<tr>
<td></td>
<td>PellPct</td>
<td>-.153</td>
<td>.008</td>
<td>-.366</td>
<td>-18.925</td>
</tr>
<tr>
<td></td>
<td>Full-time retention rate 2013 (EF2013D_RV)</td>
<td>.097</td>
<td>.014</td>
<td>.162</td>
<td>6.986</td>
</tr>
<tr>
<td></td>
<td>AdmittedPct</td>
<td>-.064</td>
<td>.005</td>
<td>-.177</td>
<td>-12.933</td>
</tr>
<tr>
<td></td>
<td>Enroll Yield</td>
<td>.037</td>
<td>.006</td>
<td>.066</td>
<td>6.657</td>
</tr>
</tbody>
</table>

a: Dependent Variable: SAT 25th
### Scenario Builder

**Example 1: Public R1 College, Low Selectivity**

#### IPEDS Reported Data 2012-13

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT 75 percentile score</td>
<td>59</td>
</tr>
<tr>
<td>6-Year graduation rate</td>
<td>56</td>
</tr>
<tr>
<td>Percent of undergraduates from out-of-state</td>
<td>30</td>
</tr>
<tr>
<td>Percentage of first-time undergraduates receiving grant or scholarship aid</td>
<td>28</td>
</tr>
<tr>
<td>Student-faculty ratio</td>
<td>13</td>
</tr>
<tr>
<td>Percent of first-time undergraduates asian</td>
<td>39</td>
</tr>
</tbody>
</table>

#### Actual versus Predicted Retention Rates

- **Predicted Retention Rate**: 81%
- **Actual Retention Rate**: 78%

![Retention Rates Graph](image-url)
**Scenario Builder**

**Example 2: Private College, URM Serving Mission**

<table>
<thead>
<tr>
<th><strong>IPEDS Reported Data 2012-13</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT 75 percentile score</td>
<td><strong>41</strong></td>
</tr>
<tr>
<td>6-Year graduation rate</td>
<td><strong>23</strong></td>
</tr>
<tr>
<td>Percent of undergraduates from out-of-state</td>
<td><strong>31</strong></td>
</tr>
<tr>
<td>Percentage of first-time undergraduates receiving grant or scholarship aid</td>
<td><strong>86</strong></td>
</tr>
<tr>
<td>Student-faculty ratio</td>
<td><strong>15</strong></td>
</tr>
<tr>
<td>Percent of first-time undergraduates asian</td>
<td><strong>0</strong></td>
</tr>
</tbody>
</table>

### Actual versus Predicted Retention Rates

- **Actual Retention Rate**: 53%
- **Predicted Retention Rate**: 50%

**Bar Chart**

![Bar Chart](chart.png)
### Scenario Builder

**Example 3: Public R1 College, Moderate Selectivity**

#### IPEDS Reported Data 2012-13

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT 75 percentile score</td>
<td>66</td>
</tr>
<tr>
<td>6-Year graduation rate</td>
<td>70</td>
</tr>
<tr>
<td>Percent of undergraduates from out-of-state</td>
<td>44</td>
</tr>
<tr>
<td>Percentage of first-time undergraduates receiving grant or scholarship aid</td>
<td>20</td>
</tr>
<tr>
<td>Student-faculty ratio</td>
<td>16</td>
</tr>
<tr>
<td>Percent of first-time undergraduates asian</td>
<td>4</td>
</tr>
</tbody>
</table>

#### Actual versus Predicted Retention Rates

- **Actual Retention Rate**: 86%
- **Predicted Retention Rate**: 87%

![Bar chart comparing actual and predicted retention rates.](chart.png)
Scenario Builder

Example 4: Private College, Highly Selective

### IPEDS Reported Data 2012-13

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT 75 percentile score</td>
<td>72</td>
</tr>
<tr>
<td>6-Year graduation rate</td>
<td>75</td>
</tr>
<tr>
<td>Percent of undergraduates from out-of-state</td>
<td>86</td>
</tr>
<tr>
<td>Percentage of first-time undergraduates receiving grant or scholarship aid</td>
<td>10</td>
</tr>
<tr>
<td>Student-faculty ratio</td>
<td>9</td>
</tr>
<tr>
<td>Percent of first-time undergraduates asian</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual versus Predicted Retention Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Retention Rate</td>
</tr>
<tr>
<td>Actual Retention Rate</td>
</tr>
<tr>
<td>0%</td>
</tr>
<tr>
<td>20%</td>
</tr>
<tr>
<td>40%</td>
</tr>
<tr>
<td>60%</td>
</tr>
<tr>
<td>80%</td>
</tr>
<tr>
<td>100%</td>
</tr>
</tbody>
</table>

Actual versus Predicted Retention Rates

- Predicted Retention Rate: 89%
- Actual Retention Rate: 85%

Actual versus Predicted Retention Rates
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
A Framework for GRD Prediction

WASC
Institutional Characteristics

WSCUC Data
Total/Graduated headcount, credits & Dropout headcount, credits

IPEDS Data
Academic Preparation
Socio-demographics
Financial aid
Institution type

Absolute Graduation Rate (Institution-type weighted)
Questions & Comments

Presentation at
http://www.uhwo.hawaii.edu/academics/oie/research-and-presentations/

Contact:

serge@unr.edu
jstanley@hawaii.edu