Predictive Modeling for Targeted Recruitment and Predicting Enrollment

Dina Bassiri, Senior Research Scientist, ACT, Inc.
Joann Moore, Research Scientist, ACT, Inc.
Overview

• What are ACT’s predictive modeling indices?
• How can the indices be used?
• Institution-specific enrollment prediction.
What is Predictive Modeling?

- Creating a statistical model to predict a future outcome
- Many approaches (traditional statistics, machine learning)
- Model development, validation, application
- Importance of choosing the right predictors
  – “All models are wrong, but some are useful” (George Box)
What are ACT’s predictive modeling indices?

Probability of enrolling…

- **Mobility**: In-state or out-of-state
- **Institution Type**: Public or private
- **Selectivity**: Open enrollment to highly selective
- **Size**: Small to large
- **Institution-Specific**: At your institution
What are ACT’s predictive modeling indices?

• Data provided when students register for the ACT
• Based on actual first year enrollment
• Available on score reports and ACT Educational Opportunity Service (EOS) files
How can the indices be used?

• Aggregate or individual student level
• Strategically target students for recruitment
• Predict whether a student will enroll at your institution
• Integrate into enrollment management system
Mobility Index

Actual Enrollment Compared to Predicted Mobility

- Enrolled In State
- Enrolled Out of State

Percent Enrolled

Mobility Index

1-5 6-20 21-35 36-55 56-65 66-80 81-99
What Factors Predict Mobility?

- Campus mix
- Preferred distance to college
- No college in mind
Institution Type Index

Actual Enrollment Compared to Predicted Institution Type

- Enrolled Private
- Enrolled Public

Percent Enrolled

Institution Type Index

1-5 6-20 21-35 36-55 56-65 66-80 81-99
What Factors Predict Institution Type?

- Campus mix
- Preferred institution type
## Selectivity Index

- **Selectivity of college or university**

<table>
<thead>
<tr>
<th>Selectivity Level</th>
<th>ACT Middle 50%</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Selective</td>
<td>25-30</td>
<td>Majority admitted from top 10% of H.S. class</td>
</tr>
<tr>
<td>Selective</td>
<td>21-26</td>
<td>Majority admitted from top 25% of H.S. class</td>
</tr>
<tr>
<td>Traditional</td>
<td>18-24</td>
<td>Majority admitted from top 50% of H.S. class</td>
</tr>
<tr>
<td>Liberal</td>
<td>17-22</td>
<td>Majority admitted from bottom 50% of H.S. class</td>
</tr>
<tr>
<td>Open</td>
<td>16-21</td>
<td>Generally open to all with H.S. diploma or equivalent</td>
</tr>
</tbody>
</table>
Selectivity Index

Actual Enrollment Compared to Predicted Selectivity

![Graph showing the relationship between Selectivity Index and Relative Probability of Enrollment for different types of institutions: Open, Liberal, Traditional, Selective, Highly Selective.](image)

- **Selectivity Index**
  - 0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

- **Relative Probability of Enrollment**

- **Institutions:**
  - Open
  - Liberal
  - Traditional
  - Selective
  - Highly Selective
What Factors Predict Selectivity?

- Campus mix
- ACT Composite Score

Pie chart showing:
- Campus Mix: 73%
- ACT Composite Score: 20%
- High School GPA: 3%
- Scores Sent to Community College: 1%
- Planned Work Hours: 1%
- Other: 2%
Institution Size Index

- **Size of college or university (full-time undergraduates)**

<table>
<thead>
<tr>
<th>Institution Size</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 5,000</td>
<td>Small</td>
</tr>
<tr>
<td>5,000 - 9,999</td>
<td>Medium</td>
</tr>
<tr>
<td>10,000 - 19,999</td>
<td>Large</td>
</tr>
<tr>
<td>20,000 and up</td>
<td>Very Large</td>
</tr>
</tbody>
</table>
Institution Size Index

Actual Enrollment Compared to Predicted Size
What Factors Predict Institution Size?

- Campus mix
- ACT Composite Score

[Diagram showing pie chart with Campus Mix at 86%, ACT Composite Score at 10%, Preferred College Size at 1%, High School Size at 1%, High School GPA at 1%, and Other at 1%]
Predicting Enrollment, Retention

• Students are more likely to attend an institution that is aligned with their preferences
• They are also less likely to transfer if good alignment
Student Preferences and Transfer Rates

Year 2 Transfer Rates by Student Institution Type Preference and Match

- Preferred 2-year
- Preferred Public 4-Year
- Preferred Private 4-Year

<table>
<thead>
<tr>
<th></th>
<th>Attended 2-year</th>
<th>Attended Public 4-Year</th>
<th>Attended Private 4-Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>11</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Preferred 2-year</td>
<td>18</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Preferred Private</td>
<td>21</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Note: The diagram shows the transfer rates for students who attended different types of institutions and preferred different types of institutions.
Student Preferences and Transfer Rates

Year 2 Transfer Rates by Location Preference and Match

- Preferred In-State College
- Preferred Out-of-State College

Attended In-State College: 13%
Attended Out-of-State College: 17%

Preferred In-State College: 15%
Preferred Out-of-State College: 13%
How can the predictive modeling indices be used?

• Target students whose preferences match your institution’s characteristics
• Use EOS search criteria related to predictive modeling indices
• Incorporate indices into enrollment prediction models
EOS Selection Criteria

- **Demographics**: gender, race/ethnicity, religion, income, parent education level
- **Geographic Region**: state, country, zip code
- **ACT Score Ranges**
- **High School Academics**: GPA, rank
- **College plans**: type, size, distance, degree aspirations, major, tuition, extracurricular activities
- **Academic ability and distance from home to campus are main factors in college choice**
# Mobility Index

At a large, public, moderately selective institution

<table>
<thead>
<tr>
<th>Mobility Index</th>
<th>Percent at Each Range (National)</th>
<th>Percent at Each Range In-State (Your Inst)</th>
<th>Percent at Each Range Out-of-State (Your Inst)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01 - 0.05</td>
<td>28</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>0.06 - 0.20</td>
<td>29</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>0.21 - 0.35</td>
<td>20</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>0.36 - 0.55</td>
<td>13</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>0.56 - 0.65</td>
<td>4</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>0.66 - 0.80</td>
<td>4</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>0.81 - 0.99</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
### Institution Type Index

<table>
<thead>
<tr>
<th>Institution Type Index</th>
<th>Percent at Each Range (National)</th>
<th>Percent at Small Private Selective</th>
<th>Percent at Large Moderately Selective</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01 - 0.05</td>
<td>16</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>0.06 - 0.20</td>
<td>48</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td>0.21 - 0.35</td>
<td>17</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>0.36 - 0.55</td>
<td>9</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>0.56 - 0.65</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>0.66 - 0.80</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>0.81 - 0.99</td>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>
## Selectivity Index

<table>
<thead>
<tr>
<th>Selectivity Index</th>
<th>Selectivity Level</th>
<th>Percent at Each Range (National)</th>
<th>Percent at Small Private Selective</th>
<th>Percent at Large Public Moderately Selective</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 1.1</td>
<td>Open</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1.2 - 2.0</td>
<td>Liberal</td>
<td>24</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2.1 - 2.8</td>
<td>Traditional</td>
<td>35</td>
<td>32</td>
<td>47</td>
</tr>
<tr>
<td>2.9 - 3.8</td>
<td>Selective</td>
<td>31</td>
<td>57</td>
<td>45</td>
</tr>
<tr>
<td>3.9 - 5.0</td>
<td>Highly Selective</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>
## Size Index

<table>
<thead>
<tr>
<th>Institution Size Index</th>
<th>Institution Size Level</th>
<th>Percent at Each Range (National)</th>
<th>Percent at Small Private Selective</th>
<th>Percent at Large Public Moderately Selective</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 - 1.6</td>
<td>Under 5,000</td>
<td>18</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>1.7 - 2.2</td>
<td>5,000 - 9,999</td>
<td>31</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>2.3 - 2.7</td>
<td>10,000 - 19,999</td>
<td>28</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>2.8 - 4.0</td>
<td>20,000 and up</td>
<td>23</td>
<td>17</td>
<td>53</td>
</tr>
</tbody>
</table>
ACT’s predictive modeling indices can be used to predict who is more likely to enroll at your institution.

Understanding where students tend to fall in various ranges of the indices at your institution can guide recruiting efforts.

Different indices may be more informative at some institutions than at others.
Overview, Institution-Specific Prediction

- Predictive Modeling Data in the ACT Electronic Record
- The Interest-Major Fit Score
- Score Senders
- What is Institution-specific Enrollment Prediction?
- What are the Input Data Files?
- What are the Outcomes?
- Summary of the model for an institution
Predictive Modeling Data in the ACT Electronic Record

- Beginning in September 2012, ACT began appending 5 new data elements to the ACT Score Report to colleges and universities:
  - Four predictive modeling indices
  - Interest-Major Fit score
- The four predictive modeling indices are also included in the ACT Educational Opportunity Service (EOS) record layout
Predictive Modeling Data in the ACT Electronic Record (cont.)

• Beginning in September 2016, the four predictive modeling indices & Interest-Major Fit scores will be included in the ACT Score Report to high schools
Value of the Student Score Report to Colleges and Universities

- Student score reports sent to colleges and universities contain data that can be used to assess:
  - student enrollment intentions
  - student level of interest in an institution
  - how well student interests, plans and needs match college offerings and strengths
Value of the Student Score Report to Colleges and Universities (cont.)

• These detailed information on student interests, plans, and needs can be used to personalize and target communications
The Interest-Major Fit Score

- Interest-major fit is derived from two data elements that are collected during ACT registration:
  - the student’s intended major chosen from a list of nearly 300 college majors
  - the student’s ACT Interest Inventory scores
The Interest-Major Fit Score (cont.)

- On the ACT Interest Inventory, students indicate whether they like, dislike, or are indifferent to 72 common activities related to four basic work tasks: data, ideas, people and things.
- Based on student responses to these items, 6 interest inventory scale scores are calculated.
- These 6 scores correspond to the 6 interest types in Holland’s theory of careers.
The Interest-Major Fit Score (cont.)

• The Interest–Major Fit score measures the strength of the relationship between:
  ➢ 6 interest inventory scale scores, and
  ➢ 6 major profile scores based on the average ACT Inventory scores across successful students within a major

• Interest-Major Fit scores range from 0 to 99
• The higher the score, the better the interest–major fit
The Interest-Major Fit Score (cont.)

- Interest-major fit clearly benefits both students and the college they attend
- Students in “good-fit majors” are more likely to stay in college, stay in their major, and finish sooner
Score Senders Are Very Productive Inquirers

- Score-sender populations apply and enroll at higher rates than almost any other inquiry type (e.g., stealth applicants, self-initiated inquiries, etc.)
- When students take the ACT, they can choose to send their scores to up to six institutions in preference order
- The more scores a student sends, the more likely they are to enroll
- Sending multiple scores shows continuous interest over time
Institution-specific Enrollment Prediction

What is institution-specific enrollment prediction?

- Creating a statistical model to predict the probability of a prospect applying and enrolling at a particular college or university
- Statistical models are updated annually
What are the Input Data Files?

- National Student Clearinghouse (CH) enrollment dataset (most recent clearinghouse file)
- High School (HS) graduate dataset (the file corresponding to the most recent clearinghouse file)
What are the Input Data Files? (Cont.)

- Library of institution information: zip code, latitude, longitude, etc. retrieved from:
  - Tables of Market Data Retrieval (MDR)
  - Institutional Data Questionnaire (IDQ)
  - College Information File (CIF)
  - College Mailing List (CML)
  - Integrated Postsecondary Education Data System (IPEDS)
What are the Outcomes?

• Every week for each participating institution an output file is generated which is:
  ➢ an exact duplicate of the file typically received from ACT each week for all students who send scores to that institution

• Appended are the following variables:
  ➢ Probability of enrollment
What are the Outcomes? (cont.)

- Mobility Index
- Institution Type Index
- Selectivity Index
- Size Index
- Percentile rank of Probability of enrollment
- Percentile rank of four indices
- Impute flag
What are the Outcomes? (cont.)

- Probability of enrollment: predicting the likelihood that a student will enroll at an institution
- Probability of enrollment is calculated by combining data in a student ACT record with weights
The weights are calculated based on selected variables in students’ ACT records and Clearinghouse enrollment data.
What are the Outcomes? (cont.)

- The probabilities of enrollment are ranked from smallest to largest

- A value of 1 indicates that the test taker is in the bottom 1% in terms of probability of enrollment

- A value of 100 indicates that the test taker is in the top 1% in terms of probability of enrollment
What are the Outcomes? (cont.)

- The percentile ranks for Enrollment, Mobility, and Institution Type are calculated based on:
  - data from all students who sent scores to an institution, and
  - who tested in October or April
- The ranks for Selectivity and Size indices are calculated across all institutions
- Percentile ranks describe the percent of students scoring at or below each index
What are the Outcomes? (cont.)

- If an important variable is missing, no probability is calculated

- Otherwise, a value is imputed for the missing variable, and the probability is calculated based on the test taker’s data and the imputed values
What are the Outcomes? (cont.)

• The impute flag of 1 refers to test takers who have missing data on one or more fields that are used for calculation of the probability of enrollment.

• In general, the imputed probabilities are very close to what would be calculated if complete data were available.
The Predictive Power of the Student Choice Set

• When students take the ACT, they can choose to send their scores to up to six institutions in preferential order

• ACT calls this mix of campuses “the choice set” or “campus mix”
The Predictive Power of the Student Choice Set (cont.)

• Through extensive research, ACT has determined that “the choice set” carries a predictive power in the models that far outweighs other variables.

• The preference order, or level of college choice, is the single most important factor in predicting that a student will enroll at a specific institution.
Summary of the Model for an Institution

• Graphical representation of the relative importance of the 8 most important variables in a pie-chart for an institution
Importance of Variables in Score Sender Model

- choice: 73%
- hsgpa: 8%
- size: 5%
- income: 1%
- work: 1%
- major: 3%
- actc: 5%
- miles: 4%
Summary of the Model for an Institution (cont.)

- A graph detailing the percentage of students who will enroll (x-axis, N), as a function of their probabilities (y-axis, Yield/Prob)

- At about 65% of Yield/Prob, 30% of students will enroll at this institution

- At about 95% of Yield/Prob, 70% of students will enroll at this institution
### Standardized Regression Weights for Score Senders Variables

- A partial table for selected variables

<table>
<thead>
<tr>
<th>Source</th>
<th>Variable</th>
<th>Weight</th>
<th>Entry Order</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>Miles</td>
<td>-0.1505</td>
<td>5</td>
<td>Examinees who live closer to your institution are more likely to enroll</td>
</tr>
<tr>
<td>ACT</td>
<td>ACT Composite</td>
<td>0.1361</td>
<td>4</td>
<td>Examinees with higher ACT scores are more likely to enroll</td>
</tr>
<tr>
<td>ACT</td>
<td>HSGPA</td>
<td>0.1900</td>
<td>3</td>
<td>Examinees with higher high school grade averages are more likely to enroll</td>
</tr>
<tr>
<td>SPS 70</td>
<td>College Size</td>
<td>0.1100</td>
<td>2</td>
<td>Examinees who prefer larger colleges are likely to enroll</td>
</tr>
<tr>
<td>ACT</td>
<td>Choice</td>
<td>-0.5683</td>
<td>1</td>
<td>Examinees who rank your institution higher as a score recipient are more likely to enroll</td>
</tr>
</tbody>
</table>
Other useful ACT research

• More information about Predictive Modeling, Interest-Major Fit, and the College Choice reports can be found at www.act.org
Research Partnership Opportunities

• We need your data to make research possible
• College Success Research
  – College GPA, Major, Degree Earned
• ELA/Writing Research
  – College Courses Taken and Grades Earned
• Please take a flyer if interested
Thank you!

• Any questions?

• Dina.Bassiri@act.org
• Joann.Moore@act.org

• For other questions contact your Account Manager