Prediction of Graduation Rates
Using First Year Indicators:
A Latent Variable Approach

Gary J. Kennedy, Ph.D.

January 22, 2019
Assumptions

1. Students in any NFYS cohort experience differential risks of not graduating within six years.
   - Importantly, this suggests that there are individual differences in students’ academic performance.
   - These differences may be captured as “mixtures” or subgroups of individuals with different hazard rates.

2. Underlying heterogeneity in graduation rates exists due to measured and unmeasured factors.
Hypotheses

• The underlying heterogeneity in graduation rates can be taken into account by latent classes of students allowing for differential effects on graduation hazard rates.

• Subgroups within a cohort can be identified based on knowledge of first-year GPA and its relationship to graduation rates
  1. First Year GPA: Intercepts and slopes from a Mixture Regression
  2. The relationship of the Mixture Regression intercepts and slopes with the graduation ‘Survival function’ intercepts and trajectories

• Mixture modeling can capture this underlying heterogeneity with a single latent categorical variable.

• This variable can then be used in subsequent cohorts to predict differential risks.
What is Mixture Modeling?

• Mixture modeling is done when a researcher hypothesizes or expects that multiple *unobserved* groups are partly responsible for the differences in the *observed* data.

• For Example
  • Student GPA is indicative of mastery of the subjects they studied.
  • *The nature of the effect of student GPA on graduation probability may be different for different groups.*
  • Reduction of hundreds of variables to a few interpretable groups as is done in market segmentation.

• “The art of unscrambling eggs.”
  DL Oberski, nd.
What is a Categorical Variable?

• An example of a *observed* (ordered) categorical variable is a Likert-scored item:

  • Thus:
    "What is your opinion of this presentation?"
    -2: Boring – a gallon of coffee with an imipramine equivalent
      would be helpful right now
    -1: Meh
    0: Adequate
    +1: Good (I just wish there was more math.)
    +2: The best presentation ever!

• A *latent* categorical variable is *discovered* within the overall heterogeneity by searching for homogeneity in patterns or clusters of observations.
  • The result is a set of discrete groups
    • High within-group homogeneity
    • High between-group heterogeneity
  • The variable can be scaled at either a nominal or ordinal level
Defining a Latent Categorical Variable:
A Discrete-time Survival Mixture Regression Model

• This method uses a specific parametric model to define the variable (i.e. “Top-down” approach)
  • Different from K-means or Hierarchical Cluster Analysis (“Bottom-up” approaches)
  • Assumes a ‘latent structure’ that can be falsified
  • Looks beyond mere similarities: does not find clusters using some arbitrary distance measure
  • Is inherently more flexible
  • Can account for measurement error

• Define the Latent Categorical Variable in terms of First Year Performance and Four-, Five-, and Six-Year Graduation Rates.

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2.@osu.edu
Assess the Latent Categorical Variable

- Assess the extent to which the categories of the latent variable that emerge:
  1. delineate the data space of the relationship between first year spring and autumn GPA and
  2. link respectively to differential gradation rate intercepts and trajectories

- Apply categories determined by the model to subsequent cohorts
  - Assess construct validity of the latent variable
    - Assess the relationship between the latent variable and other variables not included in its definition (e.g. ACT, high school class rank, demographic and socioeconomic variables, etc.)
  - Assess reliability of the latent variable (i.e. consistency) across multiple cohorts
  - Assess predictive validity of the latent variable
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Autumn 2012 NFYS Cohort

$R^2 = .383$
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Scatterplot of Spring 2013 Term GPA Regressed on Autumn 2012 Term GPA: Partitioned by Latent Class Levels
Graphical Definition of the Latent Class Categorical Variable: 
Student Academic Performance for the Autumn 2012 NFYS Cohort

<table>
<thead>
<tr>
<th></th>
<th>First Year Retention</th>
<th>Second Year Retention</th>
<th>Third Year Retention</th>
<th>Fourth Year Graduation</th>
<th>Fifth Year Graduation</th>
<th>Sixth Year Graduation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>97.1%</td>
<td>94.9%</td>
<td>94.0%</td>
<td>80.1%</td>
<td>93.9%</td>
<td>94.7%</td>
</tr>
<tr>
<td>Yellow</td>
<td>98.8%</td>
<td>98.0%</td>
<td>97.6%</td>
<td>79.6%</td>
<td>96.6%</td>
<td>98.0%</td>
</tr>
<tr>
<td>Orange</td>
<td>95.4%</td>
<td>89.6%</td>
<td>86.4%</td>
<td>46.5%</td>
<td>76.9%</td>
<td>81.9%</td>
</tr>
<tr>
<td>Red</td>
<td>42.4%</td>
<td>25.4%</td>
<td>22.9%</td>
<td>4.7%</td>
<td>9.9%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Cohort</td>
<td>92.5%</td>
<td>87.9%</td>
<td>86.0%</td>
<td>58.9%</td>
<td>80.5%</td>
<td>83.5%</td>
</tr>
</tbody>
</table>

Mahalanobis Distance

- Green/Yellow: 2.431
- Yellow/Orange: 2.448
- Orange/Red: 3.799
## Distribution of Six-year Graduation Rates Across the Latent Categorical Variable for the Autumn 2012 NFYS Cohort

<table>
<thead>
<tr>
<th>Latent Class</th>
<th>Total</th>
<th>OSU Six-year Graduation (as of 6/15/2018) (Within Group)</th>
<th>Not Graduating from OSU within Six Years (Within Cohort)</th>
<th>Degrees from Other Institutions (as of 6/29/2018)*</th>
<th>% Total Bachelors</th>
<th>% Any Degree</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>Associates</td>
<td>Bachelors</td>
<td>Certificate</td>
</tr>
<tr>
<td>Green</td>
<td>790</td>
<td>748 94.7%</td>
<td>42</td>
<td>0.6%</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Yellow</td>
<td>2,640</td>
<td>2,588 98.0%</td>
<td>52</td>
<td>0.7%</td>
<td>2</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Orange</td>
<td>3,160</td>
<td>2,579 81.6%</td>
<td>581</td>
<td>8.1%</td>
<td>40</td>
<td>152</td>
<td>3</td>
</tr>
<tr>
<td>Red</td>
<td>596</td>
<td>74 12.4%</td>
<td>522</td>
<td>7.3%</td>
<td>43</td>
<td>88</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>7,186</td>
<td>5,989 83.3%</td>
<td>1,197</td>
<td>16.7%</td>
<td>85</td>
<td>290</td>
<td>8</td>
</tr>
</tbody>
</table>

* I thank Anna Leach and Beth Venter for procuring the NSC data for this analysis.
Predicted Cohort Latent Categorical Performance Variable for Five Subsequent Cohorts Based on Autumn 2012 Results

Autumn 2013 NFYS Cohort

Autumn 2014 NFYS Cohort

Autumn 2015 NFYS Cohort

Autumn 2016 NFYS Cohort

Autumn 2017 NFYS Cohort
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:

Graduation Rates

AU12: Four-Year Graduation Rate
AU13: Four-Year Graduation Rate
AU14: Four-Year Graduation Rate
AU12: Five-Year Graduation Rate
AU13: Five-Year Graduation Rate
AU12: Six-Year Graduation Rate

Proportion Graduating

RED
ORANGE
YELLOW
GREEN
COHORT

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2@osu.edu
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts: Pre-Forgiveness Cumulative First Year GPA
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:
ACT Composite and High School Class Rank

ACT Composite

<table>
<thead>
<tr>
<th>NFYS Cohort</th>
<th>Cohort Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU12</td>
<td>23.000</td>
</tr>
<tr>
<td>AU13</td>
<td>26.000</td>
</tr>
<tr>
<td>AU14</td>
<td>27.000</td>
</tr>
<tr>
<td>AU15</td>
<td>28.000</td>
</tr>
<tr>
<td>AU16</td>
<td>29.000</td>
</tr>
<tr>
<td>AU17</td>
<td>30.000</td>
</tr>
</tbody>
</table>

High School Class Rank

<table>
<thead>
<tr>
<th>NFYS Cohort</th>
<th>Cohort Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU12</td>
<td>75.000</td>
</tr>
<tr>
<td>AU13</td>
<td>80.000</td>
</tr>
<tr>
<td>AU14</td>
<td>85.000</td>
</tr>
<tr>
<td>AU15</td>
<td>90.000</td>
</tr>
<tr>
<td>AU16</td>
<td>95.000</td>
</tr>
<tr>
<td>AU17</td>
<td>100.000</td>
</tr>
</tbody>
</table>
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:
Proportion of STEM Majors
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:
Proportion of Under-represented Minority Students

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2.@osu.edu
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:
Proportion of First Generation Students

Proportion First Generation

NFYS Cohort

AU12  AU13  AU14  AU15  AU16  AU17

RED  ORANGE  YELLOW  GREEN  Cohort Mean
Comparison of Autumn 2012 with the 2013 through 2017 NFYS Cohorts:
Proportion of Pell Eligible Students

Proportion Receiving Pell

NFYS Cohort

AU12  AU13  AU14  AU15  AU16  AU17

Cohort Mean

RED
ORANGE
YELLOW
GREEN

Contact: kennedy2.@osu.edu
Comparison of Autumn 2012, 2013, and 2016 NFYS Cohorts: Selected SES Census Variables – Percent in Block Group with a College Degree

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2.@osu.edu
Comparison of Autumn 2012, 2013, and 2016 NFYS Cohorts: Selected SES Census Variables – Percent in Block Group with No High School Diploma

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2.@osu.edu
Comparison of Autumn 2012, 2013, and 2016 NFYS Cohorts:
Selected SES Census Variables – Percent in Block Group with Never Married

THE OHIO STATE UNIVERSITY
Enrollment Services – Analysis and Reporting

Contact: kennedy2.@osu.edu
Comparison of Autumn 2012, 2013, and 2016 NFYS Cohorts:
Selected SES Census Variables – Percent in Block Group Divorced

Proportion in Block Group Divorced

NFYS Cohort

AU12 AU13 AU14 AU15 AU16 AU17

RED ORANGE YELLOW GREEN Cohort Mean
Relationship between the Student Academic Performance Latent Variable and Two Motivational Variables for the Autumn 2016 and 2017 NFYS Cohorts
Relationship between the Student Academic Performance Latent Variable and Two Academic Advising Variables

![Graph showing relationship between student academic performance and academic advising variables.](image-url)
Conclusions

• The latent categorical variable
  • Models non-normally distributed academic performance distributions
  • Provides statistically sound categorization
  • Appears to show some consistency over cohorts
  • May be a valid measure of student academic performance
  • Shows some promise in predicting graduation rates

• Limitations
  • Takes a long time (six-years) to define the variable
  • It is still unclear whether the patterns seen in 2012 cohort will apply to the 2018 cohort
  • Causal relationships are unclear
Thank You!

Questions, Comment, Concerns?