Modeling First-Year Retention at Western Michigan University

Office of Institutional Research | Western Michigan University
January 28, 2019
Talk Outline

1. Data on Retention at WMU
2. The Model: what are we trying to do?
3. Results: which predictors are associated with student retention?
4. Predictions: can we predict the next year's retention?
5. Extensions: how do we use and expand the model going forward?
Definitions

FTIAC: Full-time, first-time-in-college students

Retention: percentage of FTIACs that return to WMU for their second fall semester (i.e. are registered for classes as of census)

Probability of Retention: the probability that an individual student will be retained. Mathematically we often report a number between 0 and 1. I will report as a percentage between 0 and 100 (e.g. 60% probability of being retained).

The Office of Institutional Research collects and curates data each fall in order to report retention statistics to the Integrated Postsecondary Education Data System (IPEDS) and the Consortium on Student Retention Data Exchange (CSRDE).

Retention has increased by 3.5 percentage points since 2006.

…and we have plateaued at a higher level in the last three years.
How do we compare to other schools?

Source: IPEDS
How do we compare to other schools?
How do we compare to other schools?

Given the ACT scores of our incoming FTIACs, our retention is about 3-7% higher than expected relative to our peers in 2015.
How do we compare to other schools?

The percent of students we admitted in 2015 tells a similar story (about 1-3% better than expected relative to our peers).
Average Retention by Ethnicity (2011-2015)

- International (n = 232)
- Asian (n = 244)
- White (n = 10810)
- Native Hawaiian or Other Pacific Islander (n = 20)
- Hispanic (n = 915)
- No Response (n = 101)
- Two or More Races (n = 554)
- Black or African American (n = 2269)
- American Indian or Alaska Native (n = 50)

Overall WMU Average
Average Retention by College (2011-2015)

Note: 'Other' are all University Curriculum (INTO) students
Average Retention by Department (2011-2015)
Average Retention by Residency (2011-2015)

- International, Non-Resident (n = 232)
- Domestic, Non-Resident (n = 1665)
- Resident (n = 13298)

Overall WMU Average
In Summary

WMU retention is about 3 percentage points higher than it was a decade ago.

Given our selectivity, our raw retention rates have been a bit higher than expected relative to our peers (depending on your metric).

There is a lot of variation in the raw retention rates among different groups of students on campus.
Why Statistical Modeling
The raw data do not tell the whole story..
Why Statistical Modeling (a hypothetical example)

International, non-resident students retain about 10% better than residents, and domestic, non-residents about 5% better…
Why Statistical Modeling (a hypothetical example)

Maybe the non-resident students we admit are more academically prepared for college...

![Graph showing statistical modeling example]
Why Statistical Modeling (a hypothetical example)

...and the underlying retention rates would look something like this if we controlled for preparedness.
The Model
The Purposes of Building a Model

To **understand** which factors are most strongly associated with student retention.

To **predict** retention for existing cohorts or sub-groups already on campus.

To **simulate** hypothetical changes in retention rates based on proposed changes in university policy, procedures, and programs.
The Model

Uses a hierarchical, logistic regression structure to predict the probability of retention of individual students based on their individual characteristics (i.e. predictors):

\[
\text{Probability that a student returns} = \text{predictor 1} + \text{predictor 2} + \text{predictor 3} + \text{predictor 4} + \text{predictor 5} + \ldots
\]

It's a bit more complicated than that...but not that much more
The Model

Combines retention data on students from the previous five cohorts (2011-2015)
Student-level predictors currently in the model

- ethnicity
- college and department
- county
- residency (international, non-resident, resident)
- high school GPA (HSGPA) and number of AP courses
- expected family contribution (EFC):
  - students who need 'full' aid (EFC = 0)
  - students who need 'some' aid ($1 < EFC < $100,000, most between $1,000 and $100,000)
  - students who did not submit a FAFSA ('no' aid)
- expected family contribution ($) of students ('some' group only)
Predictors purposely not in the model

A few predictors were considered but left out of the model:

- ACT score
- Honors college affiliation
- High school
Results
Continuous Predictors

We have three student-level, continuous predictors in our model:

- high school GPA (HSGPA)
- number of AP courses
- expected family contribution ($) of students ('some' group only)
Marginal Effects: Continuous Predictors

Our model estimates the difference in the probability that a student will return between an average student and a student with a predictor value 1 standard deviation above (or below) the mean:
Between 2011 and 2015, a student with a HSGPA of 3.31 would have an average retention probability of 76%.
Marginal Effects: High School GPA

A student with a HSGPA that was 1 SD higher (about 3.81)...

![Graph showing marginal probability of retention against high school GPA]
Marginal Effects: High School GPA

... would have a 7 percentage point higher probability of being retained.
Marginal Effects: Number of AP courses

Similarly, a student that took one additional AP course more than the average student will have a 3 percentage point higher probability of being retained.
Marginal Effects: Expected Family Contribution

A student with an EFC about 4.5 times greater than average will have a 2 percentage point higher probability of being retained.
Marginal Effects: Continuous predictors

Comparing marginal effects

- High school GPA
- Number of AP courses
- log2(Expected family contribution)

marginal change in retention
Group Predictors

We have six group predictors in our model:

- ethnicity
- college
- department
- county
- residency (international, non-resident, resident)
- expected family contribution (EFC):
  - students who need 'full' aid (EFC = 0)
  - students who need 'some' aid ($1 < EFC < $100,000, most between $1,000 and $100,000)
  - students who did not submit a FAFSA ('no' aid)
Marginal Effects: Group Predictors

Which group predictors are most important for predicting retention?

- department
- county
- family contribution
- college
- residency
- ethnicity

Variance in retention among levels within each group
Marginal Effects: Department

![Graph showing marginal effects by department]
Marginal Effects: County

Looking at the top 10 and bottom 10 counties in by marginal effect:

**Upper 10**
- Monroe, MI (n = 89)
- McHenry, IL (n = 58)
- Saginaw, MI (n = 108)
- Montcalm, MI (n = 37)
- Dickinson, MI (n = 18)
- Wayne, MI (n = 1742)
- Livingston, MI (n = 372)
- Lenawee, MI (n = 154)
- Mecosta, MI (n = 16)
- Washtenaw, MI (n = 514)

**Lower 10**
- St Joseph, IN (n = 31)
- Macomb, MI (n = 925)
- Jackson, MI (n = 266)
- Muskegon, MI (n = 206)
- Newaygo, MI (n = 39)
- Eaton, MI (n = 160)
- Calhoun, MI (n = 397)
- Van Buren, MI (n = 417)
- Ottawa, MI (n = 453)
- Berrien, MI (n = 306)
- Grand Traverse, MI (n = 59)

marginal effect on retention %
Marginal Effects: County

counties with at least 15 students

median marginal effect on retention

+3%
0%
-3%
Marginal Effects: Socioeconomic Status

Both the 'Some' and 'None' groups retain about 4 to 8 percentage points (on average) better than the 'Full' group.
Marginal Effects: College

- Fine Arts
- Health & Human Services
- Aviation
- Education & Human Development
- Haworth College of Business
- Other
- Arts & Sciences
- Engineering & Applied Sciences

marginal effect on retention

-15% -10% -5% 0 +5% +10% +15%
Marginal Effects: Residency Status

Compared directly, both international and domestic non-residents retain better than our resident students, (though the differences may be small).
Marginal Effects: Ethnicity

There is practically no difference between ethnicities in their associated marginal retentions.
Marginal Effects: Ethnicity

There is practically no difference between ethnicities in their associated marginal effects on retention.
Marginal Effects: Summary

- Academic preparedness (i.e. high school GPA and No. AP courses) is the strongest predictor of individual retention at WMU.
- Students whose expected family contribution is zero retain at substantially lower rates than other students, and the more students can contribute financially, the more likely they are to return.
- Non-resident (both international and domestic) may retain at higher rates than residents students.
- There is a substantial amount of variation in retention associated with home counties, departments, and to a lesser extent colleges.
- Student ethnicity is associated with little to no additional variation in retention.
- None of these effects can be considered directly causal!
Predicting Future Retention
Predicting Future Retention

We can use the fitted model to predict the 2016 cohort's retention rate!
Predicting Future Retention

Remember…

\[
\text{Probability that a student returns} = \text{predictor 1} + \text{predictor 2} + \ldots
\]
Predicting Future Retention

Predicted retention for 2016.
Predicting Future Retention

Zooming in...
Predicting Future Retention

...and adding the actual retention for the 2016 cohort.
Final Thoughts
Final Thoughts

1. Understanding and predicting a binary outcome like first-year retention is challenging!

   - We only ever measure each student once!

   - A student with a 1% 'probability' and 99% 'probability' of retaining can both return for their second year!

   - We can never measure probability directly, only zeros and ones

   - Even with a lot of data our estimates may be accurate but have low precision (i.e. high uncertainty)
Final Thoughts

1. Understanding and predicting a binary outcome like first-year retention is challenging!

2. There are many reasons why students return for their second year
   - any one factor might be really important for any given student but have poor predictive ability overall (think personal reasons)
   - not returning to WMU means different things for different students
   - all those little things collectively add up to a single overall FTIAC retention
Final Thoughts

1. Understanding and predicting a binary outcome like first-year retention is challenging!
2. There are many reasons why students return for their second year
3. There is a lot variance in retention that can be explained 'statistically' without providing much understanding in reality
   - some predictors closely measure what we think they measure (i.e. HSGPA)
   - others are just placeholders for multiple underlying factors (e.g. variation among departments or counties)