

# Stopouts and Readmits: Using Student Record and NSC Data to Predict Re- Enrollment

Jerret K. LeMay,  
Research Analyst,  
Institutional Research & Planning  
Binghamton University

Funded through the State of New York / UUP Professional Development  
Committee

# Why?

- Stopout Report
- Gap in Enrollment Projection Model (EPM)
  - Enrollment Estimates by:
    - UG/GD,
    - school,
    - New / Continuing / Returning
  - Feeds Housing Estimates, Tuition & Fees Projections

# Literature

- Horn (1998) - majority of stopouts (64%) return to higher education, and 42% of those re-enroll at the same institution they left. 27% of stopouts return to their original institution
- Adelman (1999) - alternating and simultaneous enrollment patterns, “portfolio building,” where the majority (61%) of those who attended two schools returned to the first.

# Goal

- Predict, at the aggregate level, the number of readmitted students for a given semester using student record data and data from the National Student Clearinghouse (NSC)

# Data Submitted to NSC

- 12,654 records of degree-seeking and non-degree undergraduate stopouts
  - “stopout” = did not receive degree & did not register for next major semester
  - Fall 1992 until Fall 2003
  - instances, not necessarily people

# What came back

- 17,254 records
- Business decisions
  - Dates vs. Years/semesters
  - Trumps (co-enrolled, FT/PT/Unknown etc.)
  - Date of *initial* enrollment
  - Merge back into student record dataset

# Data Structure

- Person-period dataset - each student had a number of separate records equal to the number of semesters under consideration. Student record information existed for semesters in which the students were enrolled, and blank records were inserted for subsequent semesters.
- Merged in NSC data
- “ugread” – dependent variable - readmitted as an undergraduate student in the next (major) semester
- Springs only (predicting fall readmits)

# Before the Regression...

- Descriptives – Example: BU was institution of first choice for 49% of returning students, compared to 27% of non-returning students
- Bivariate – like frequencies, # of semesters missed, financial aid info, type of transfer institution (if they did transfer), and whether BU was the institution of first choice

# Multivariate Model

- Logistic Regression
- 3 Criteria:
  - Variable significant
  - Variable improved the fit ( $-2 \log$  likelihood) and quality (c) of the model
  - Improved Predictive accuracy (aggregate)

# Significance

## Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi Sq
Intercept	1	-4.2566	0.0834	2607.1542	<.0001
Missed 1 semester	1	1.6903	0.0902	351.4925	<.0001
Medium need	1	0.2232	0.1073	4.3250	0.0376
Trans NY 2yr	1	0.9225	0.0935	97.2525	<.0001
Missed 10+	1	-2.5633	0.2478	106.9605	<.0001
2yrs at BU	1	0.6296	0.1080	33.9980	<.0001
Missed 7-10	1	-1.8597	0.2073	80.5039	<.0001
2nd sem senior	1	-0.6909	0.1444	22.8928	<.0001
PLUS Loan amount	1	0.000046	0.000028	2.7773	0.0956
BU first choice	1	0.1971	0.1025	3.6973	0.0545

# Fit and Quality

## Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	6169.684	4901.798
SC	6178.277	4987.729
-2 Log L	6167.684	4881.798

## Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.6	Somers' D	0.731
Percent Discordant	11.5	Gamma	0.761
Percent Tied	4.0	Tau-a	0.021
Pairs	23284145	c	0.865

# Predictive Accuracy

Spring	ugread	predicted	diff	abs_diff
1999	85	100.749	15.7491	15.7491
2000	122	113.002	-8.9982	8.9982
2001	140	127.339	-12.6606	12.6606
2002	131	124.996	-6.0044	6.0044
2003	115	126.915	11.9146	11.9146
				=====
				55.3269

Intercept-only model = 77

# How Did It Do?

<b>Fall 2004</b>	<b>Prediction</b>	<b>Actual</b>	<b>Difference</b>
New method	156	169	-13
Old methods:			
Previous year	134	169	-35
3 year average	147	169	-22
5 year average	140	169	-29

# Other Significant Variables?

- Many significant variables which also made for better model, but didn't improve aggregate predictive accuracy
- Variation from year to year (significance, parameter estimates, standard error)

# So?

- NSC Enrollment Search
  - Despite limitations
- Person-Period Dataset
  - Incorporation of NSC data
  - “Carry forward”
  - Flexibility
- Regression – three criteria

# Contact Information

Jerret K. LeMay

[jlemay@binghamton.edu](mailto:jlemay@binghamton.edu)

Institutional Research & Planning

<http://buoir.binghamton.edu>

(Paper & presentation are available here)

Binghamton University

<http://www.binghamton.edu>