Identifying factors influencing timely graduation: Recent research and developments

Yusuke Kuroki

Sunny Moon

November 9th, 2017

Graduation Initiative 2025 Goals
CSU Fullerton

Metric	2025 Goal	Most Recent Rate
Freshman 6-Year Graduation	**75%**	62%
Freshman 4-Year Graduation	**44%**	22%
Transfer 2-Year Graduation	**44%**	32%
Transfer 4-Year Graduation	**85%**	75%
Gap - Underrepresented Minority	**0**	9 % points
Gap – Pell	**0**	4 % points

CALIFORNIA STATE UNIVERSITY FULLERTON

"Well, I already knew that": Can you go beyond the quantitative findings on admissions available graduation rate variables?

Based on Graduation Rates of First-Time Full-Time Freshmen and Full-/Part-time Transfer Students

November 18, 2010

Sunny Moon
James Hershey
Ed Sullivan

Institutional Research and Analytical Studies
California State University, Fullerton

Results from 7-yrs ago…

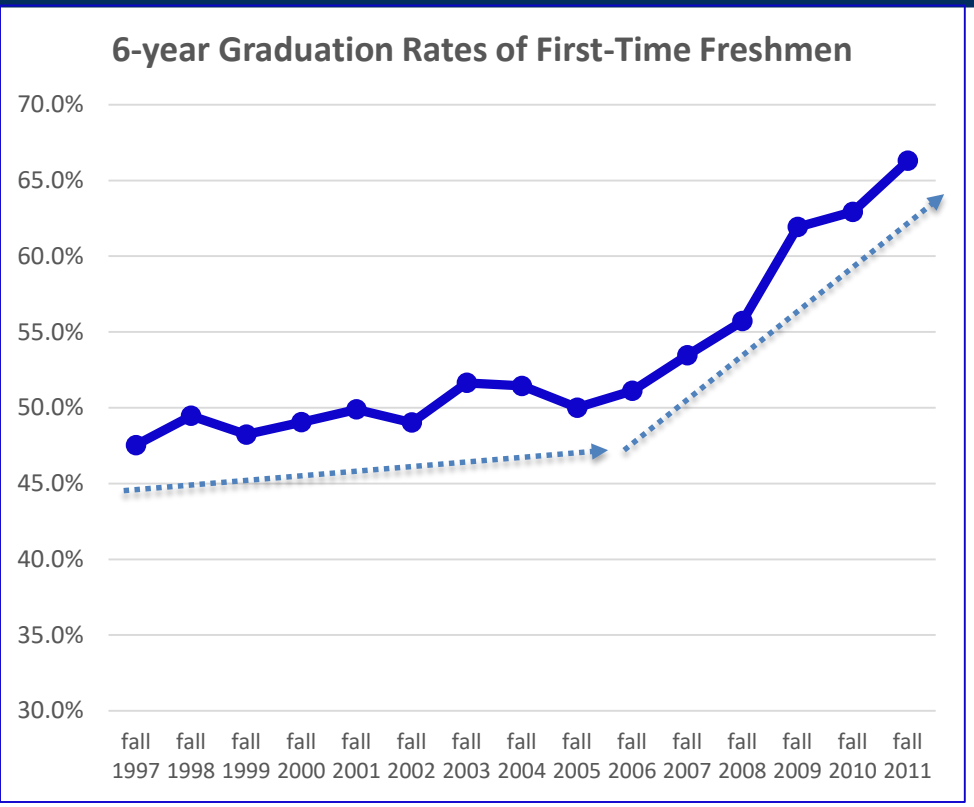
- Predicting 6-yr graduation of Full-time FTF (cohorts from fall 1998 to 2002)
  - 13,412 FTF students entered
  - 6,589 (49.1%) graduated in six years or less
- Predictors included:
  - HS GPA, SAT, ELM
  - Gender
  - Ethnicity
  - Major in 1st term
  - Admissions area
  - Units taken in 1st Term

Results from 7-yrs ago…

- Predicting 4-yr graduation of Transfers (cohorts from fall 1998 to 2004)
  - 33,839 transfer students entered
  - 21,778 (64%) graduated in four years or less
- Predictors included:
  - Transfer GPA
  - Major
  - Full-time status
  - Ethnicity
  - Gender

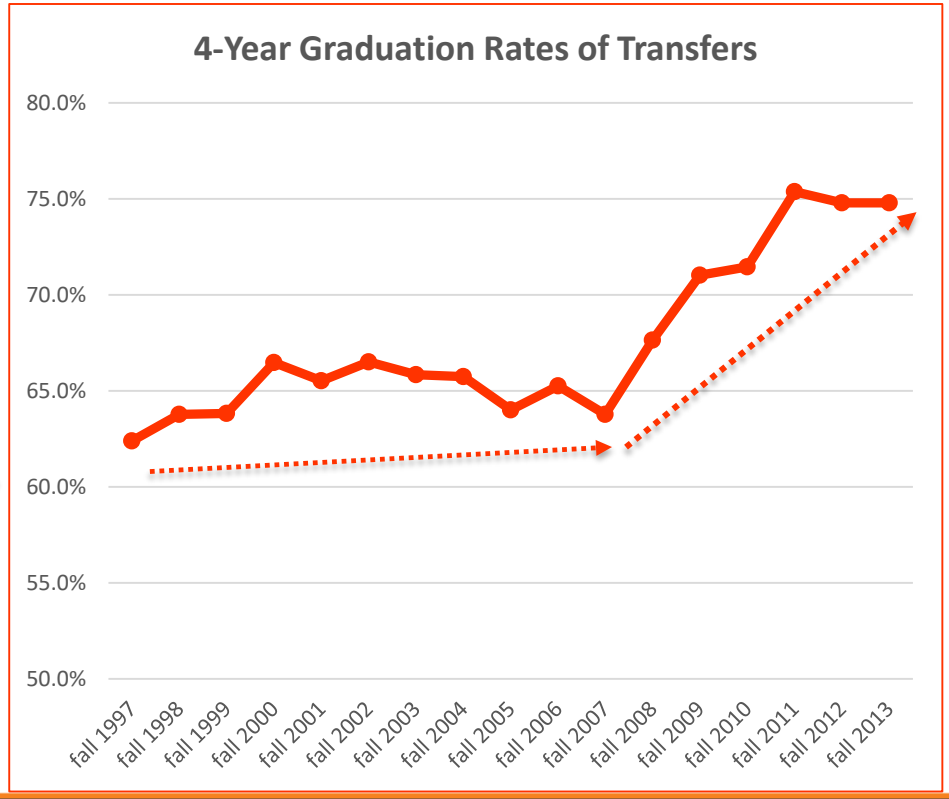
Summary

- Less well academically prepared students are less likely to succeed.
- Men are less likely to succeed than women
- URM are less likely to graduate
- Higher # of Units in 1st term is associated with graduation
- STEM majors less likely to succeed
- Students from outside of our local admissions area are less likely to succeed
- Freshmen who complete remediation in mathematics early in their academic careers are more likely to succeed.



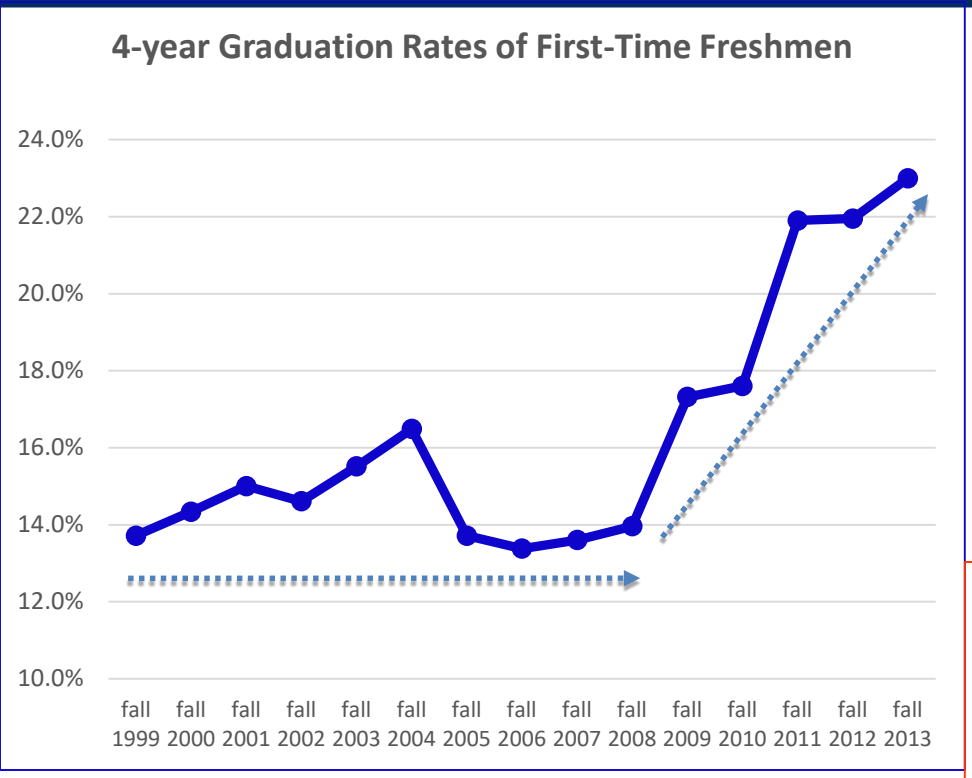
6-Year Graduation Rates

First-Time Freshmen



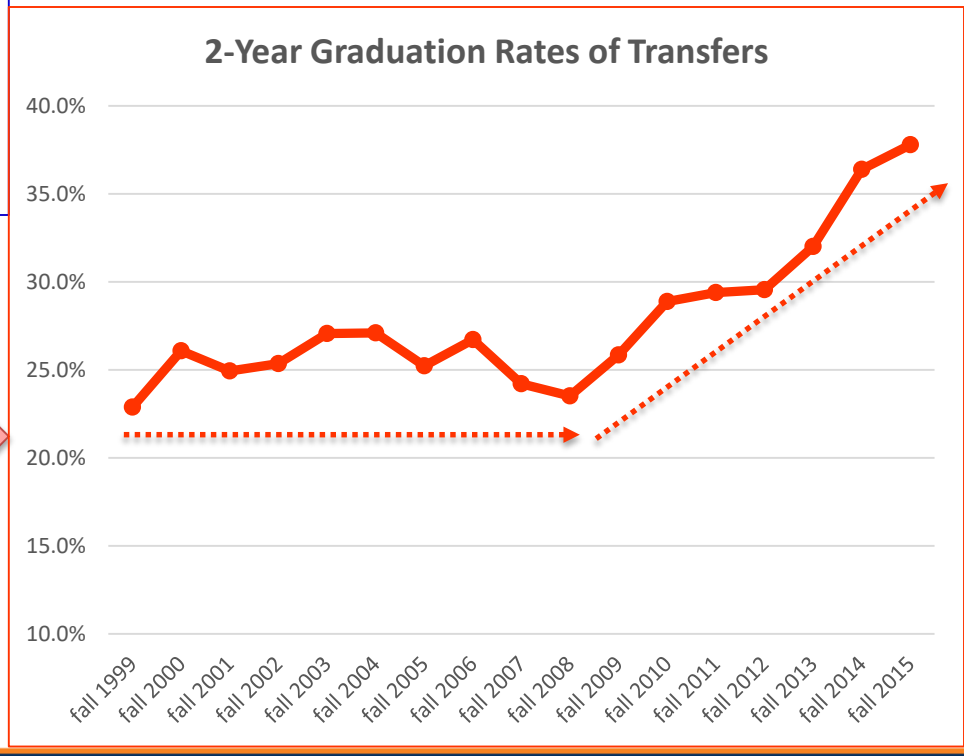
4-Year Graduation Rates

Transfers



4-Year Graduation Rates

First-Time Freshmen



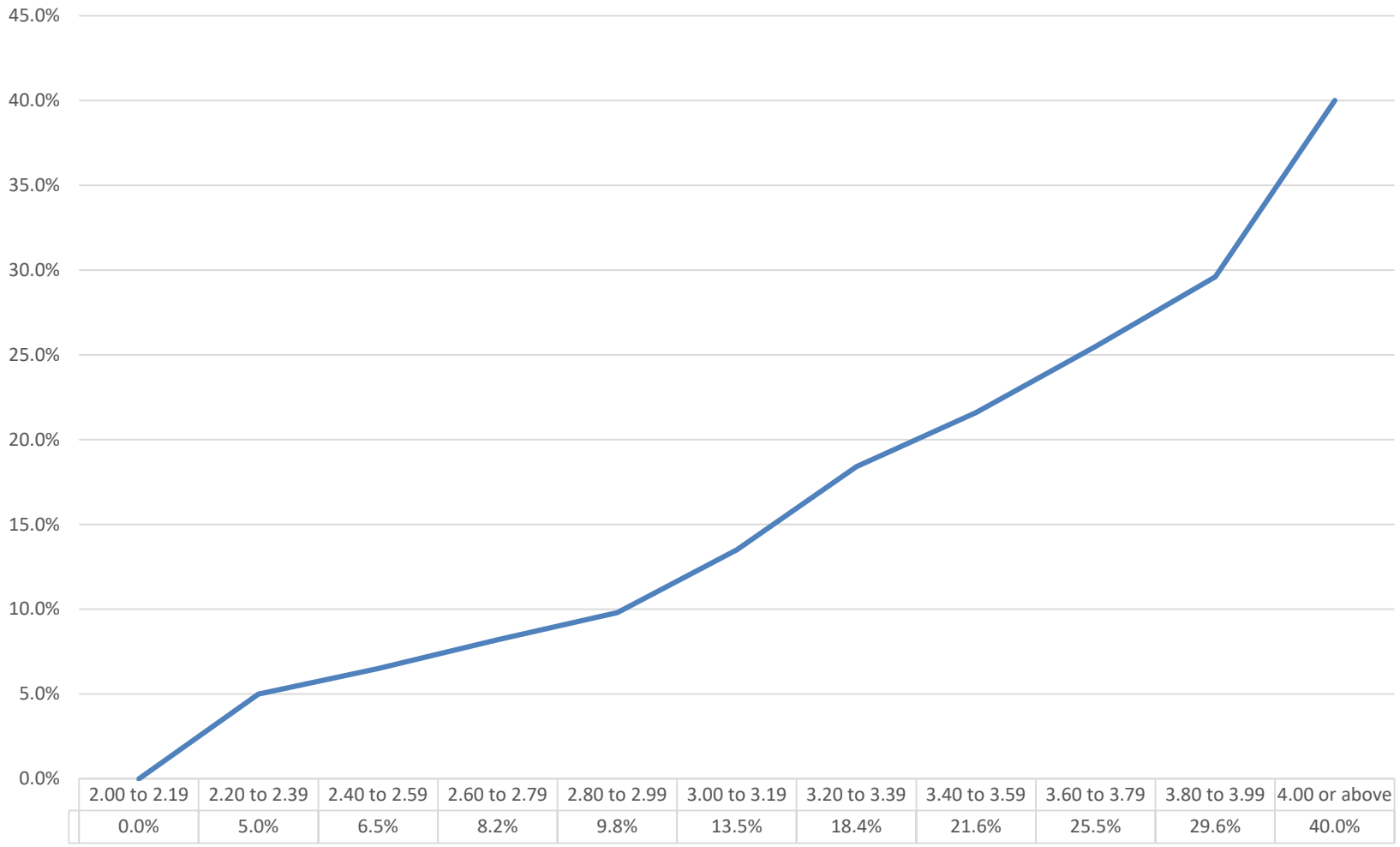
2-Year Graduation Rates

Transfers

Four-year graduation rates by student characteristics

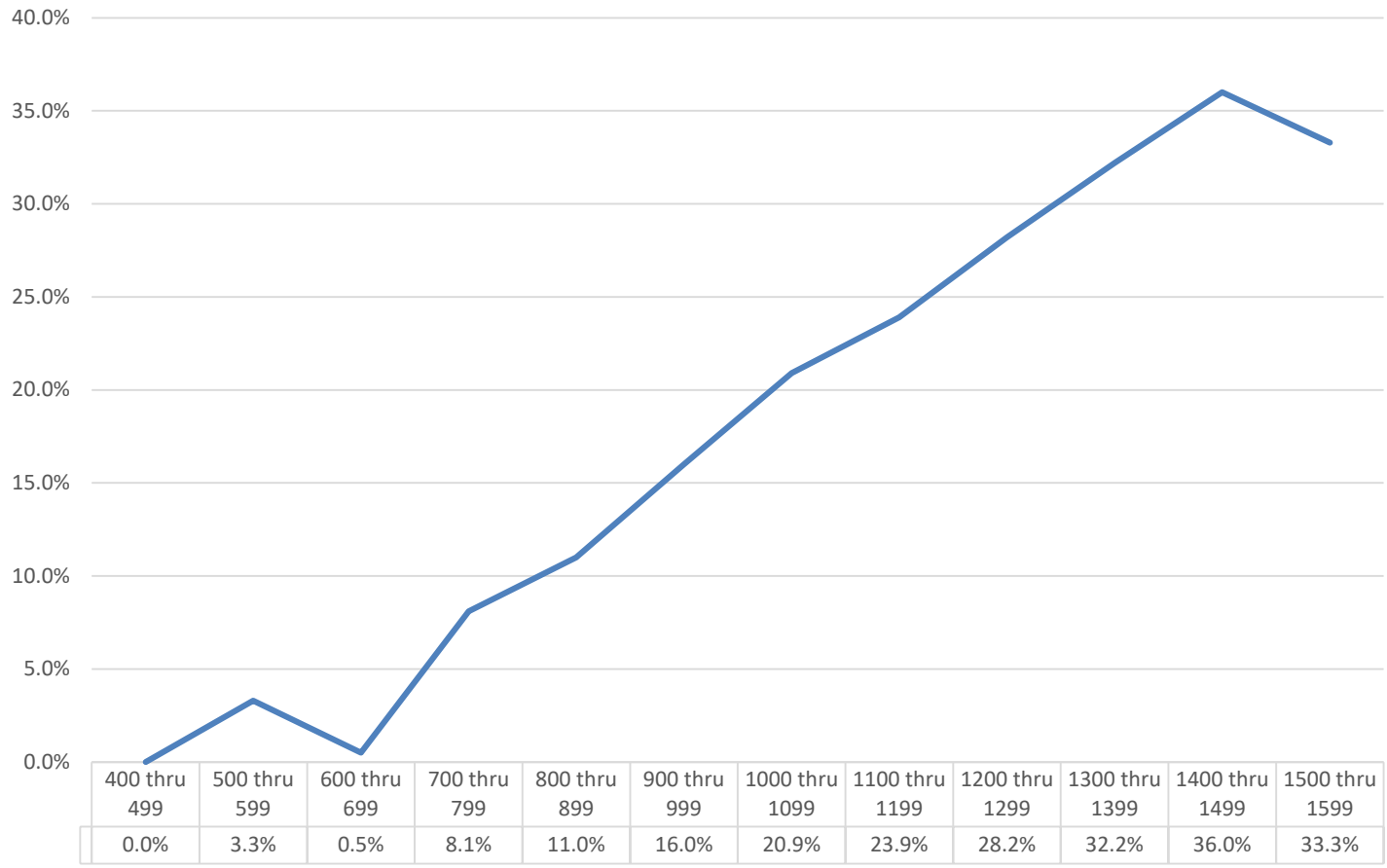
- Data
  - FTF from fall 2008 to fall 2012 cohorts
    - 20621 Full-time FTFs.
    - 3823 (18.5%) graduated in 4 years or less.

4-yr graduation rate by HS GPA



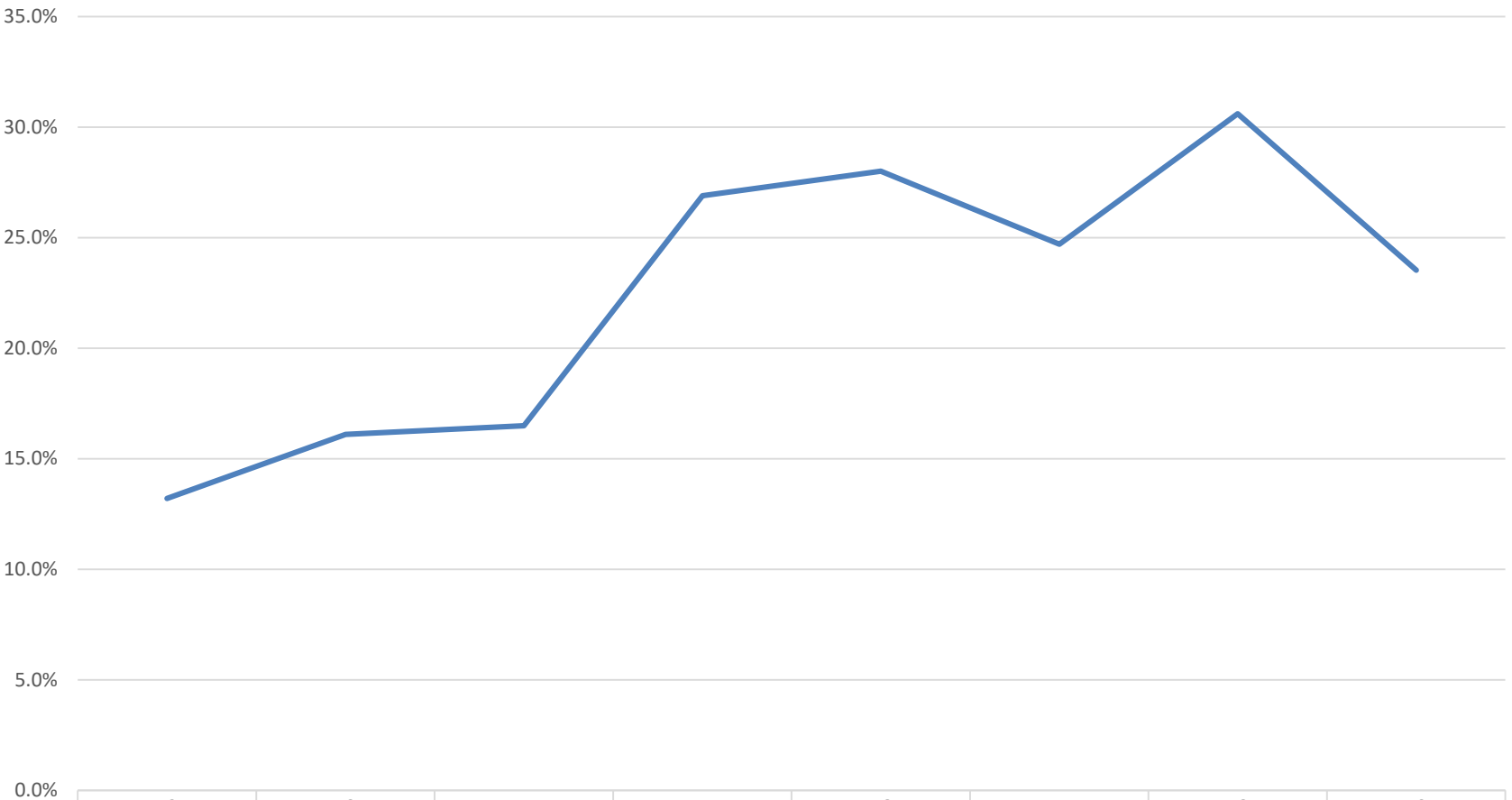
2.00 to 2.19	2.20 to 2.39	2.40 to 2.59	2.60 to 2.79	2.80 to 2.99	3.00 to 3.19	3.20 to 3.39	3.40 to 3.59	3.60 to 3.79	3.80 to 3.99	4.00 or above
0.0%	5.0%	6.5%	8.2%	9.8%	13.5%	18.4%	21.6%	25.5%	29.6%	40.0%

4-yr graduation rate by SAT



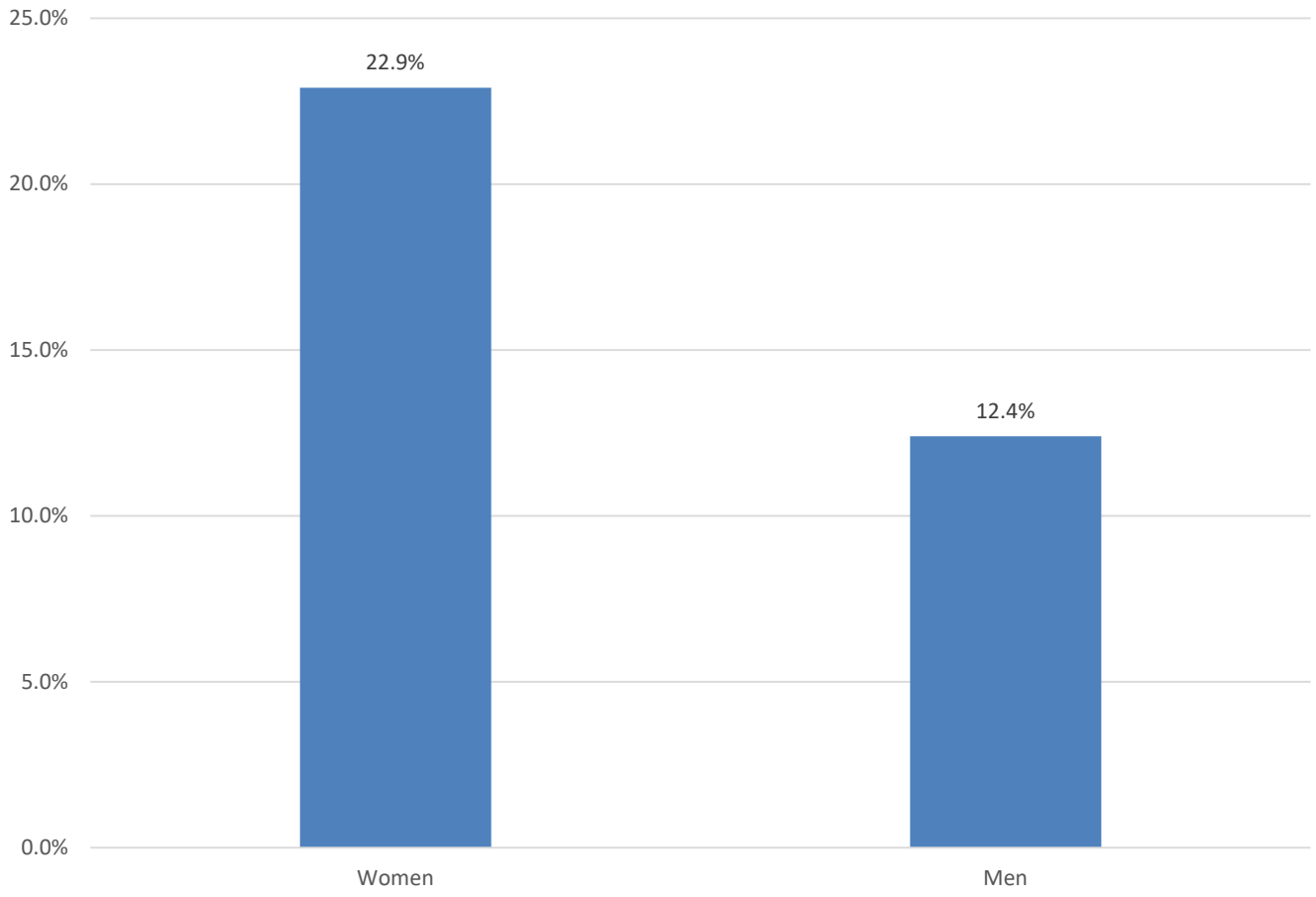
400 thru 499	500 thru 599	600 thru 699	700 thru 799	800 thru 899	900 thru 999	1000 thru 1099	1100 thru 1199	1200 thru 1299	1300 thru 1399	1400 thru 1499	1500 thru 1599
0.0%	3.3%	0.5%	8.1%	11.0%	16.0%	20.9%	23.9%	28.2%	32.2%	36.0%	33.3%

4-yr graduation rate by 1st Term Units

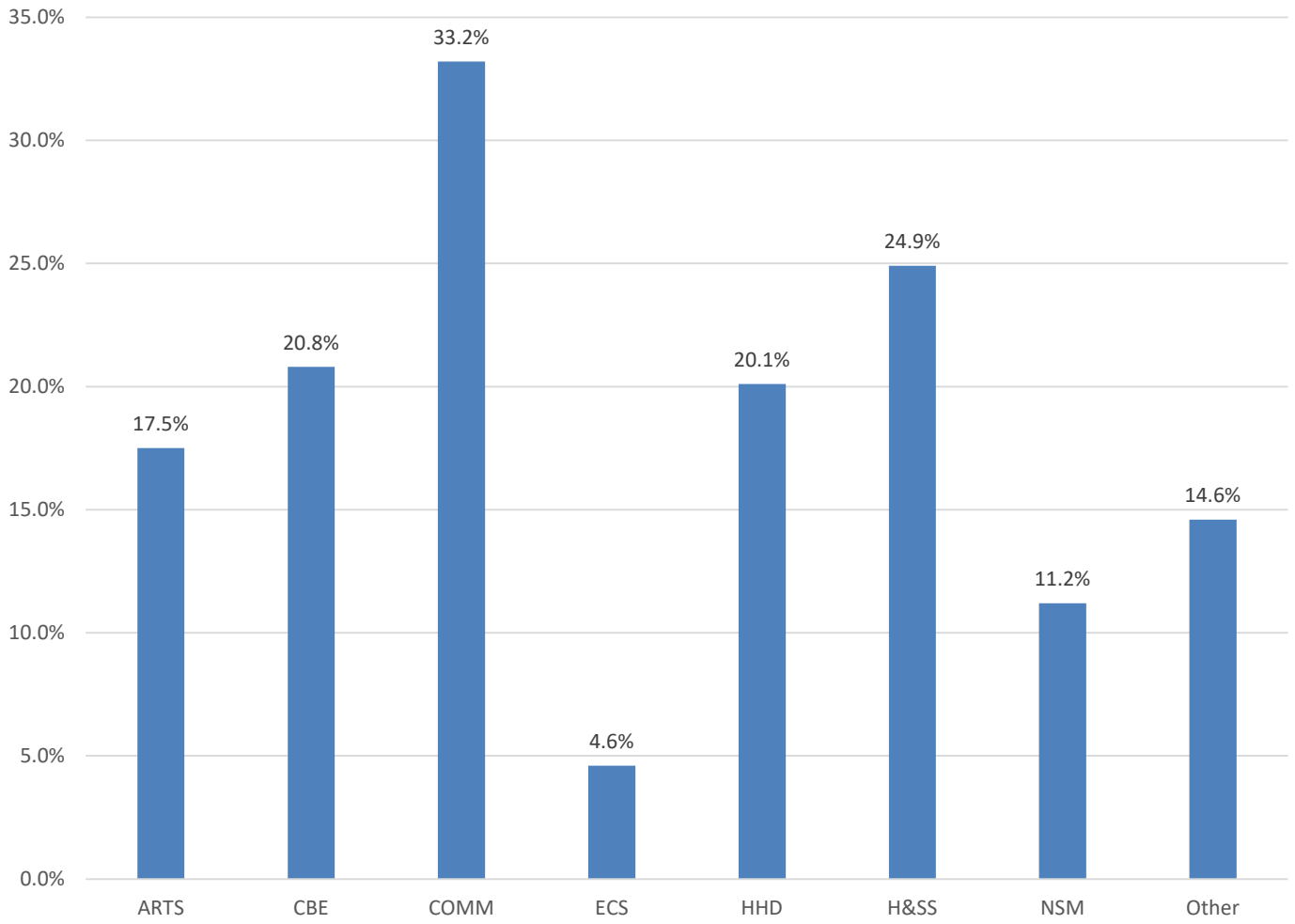


12	13	14	15	16	17	18	19
13.2%	16.1%	16.5%	26.9%	28.0%	24.7%	30.6%	23.5%

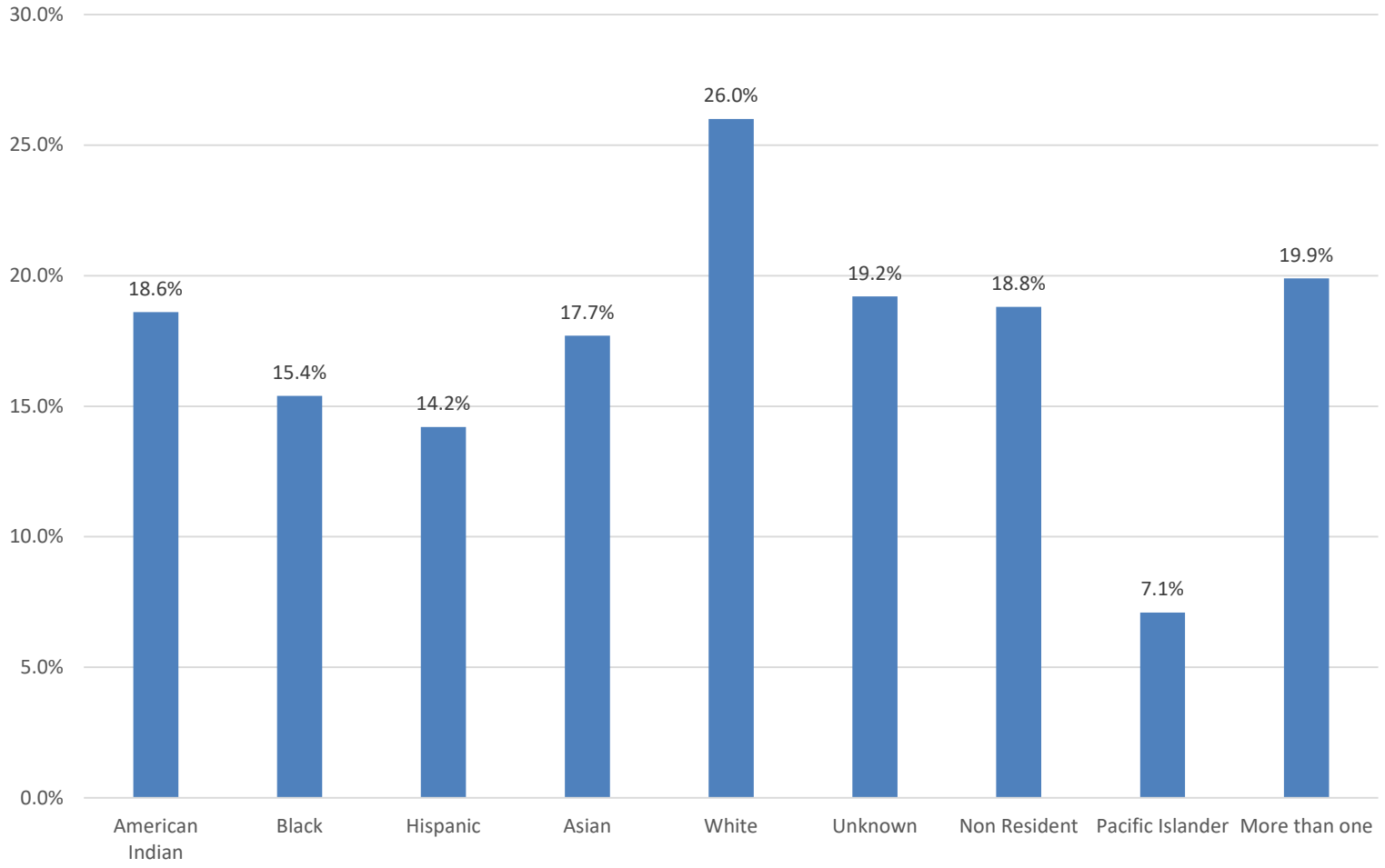
4-yr graduation rate by gender



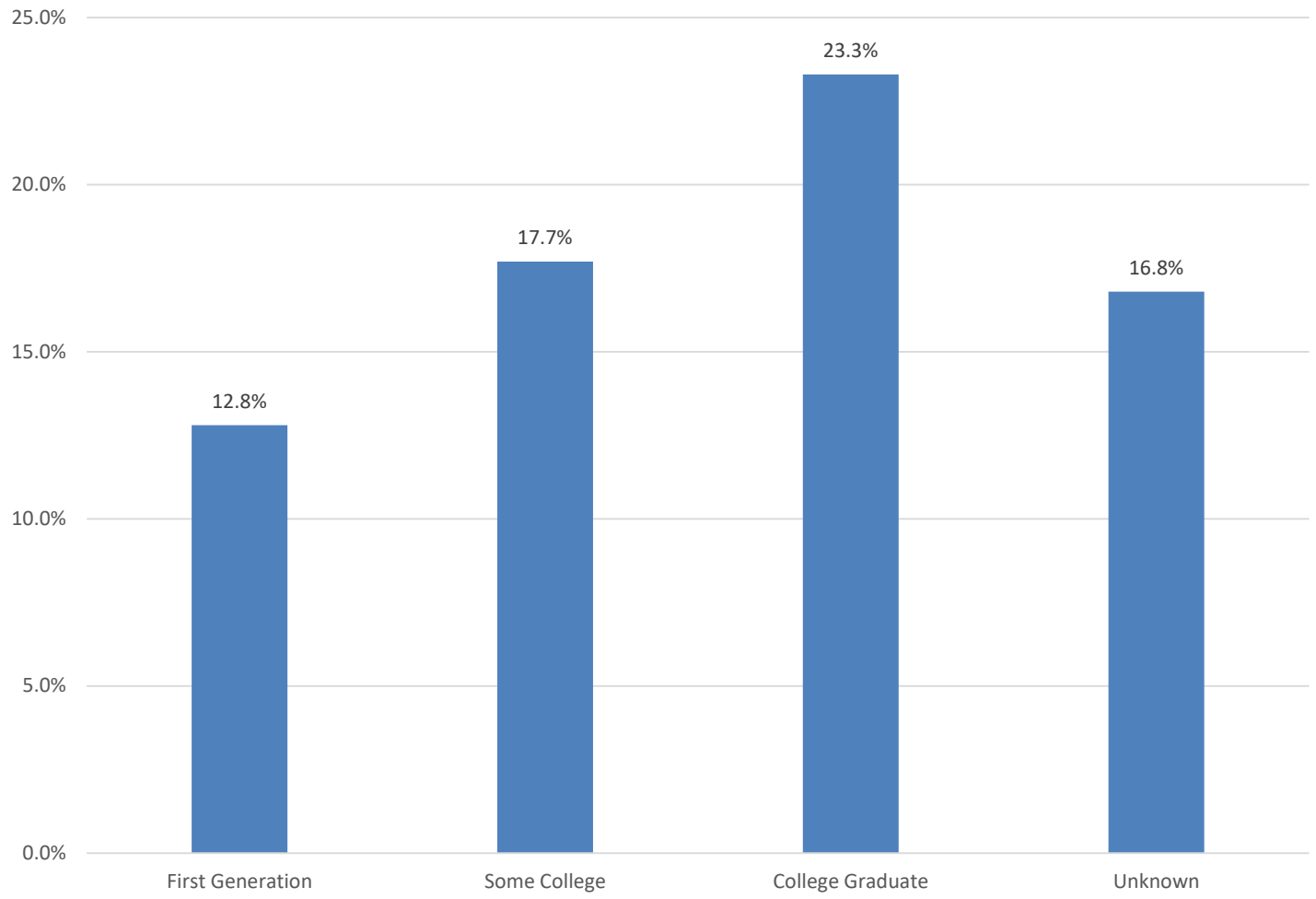
4-yr graduation rate by Major at Entry



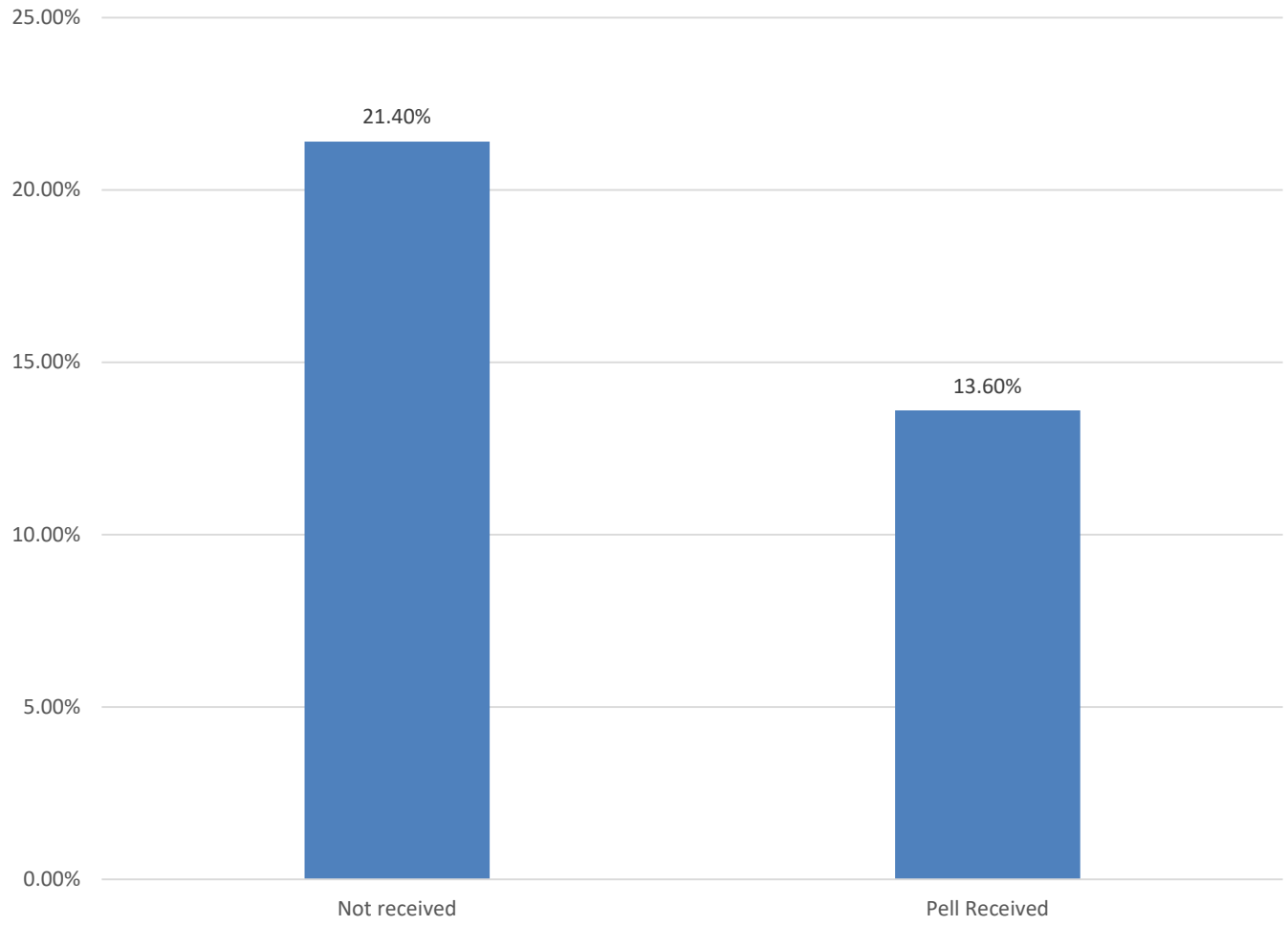
4-yr graduation rate by ethnicity



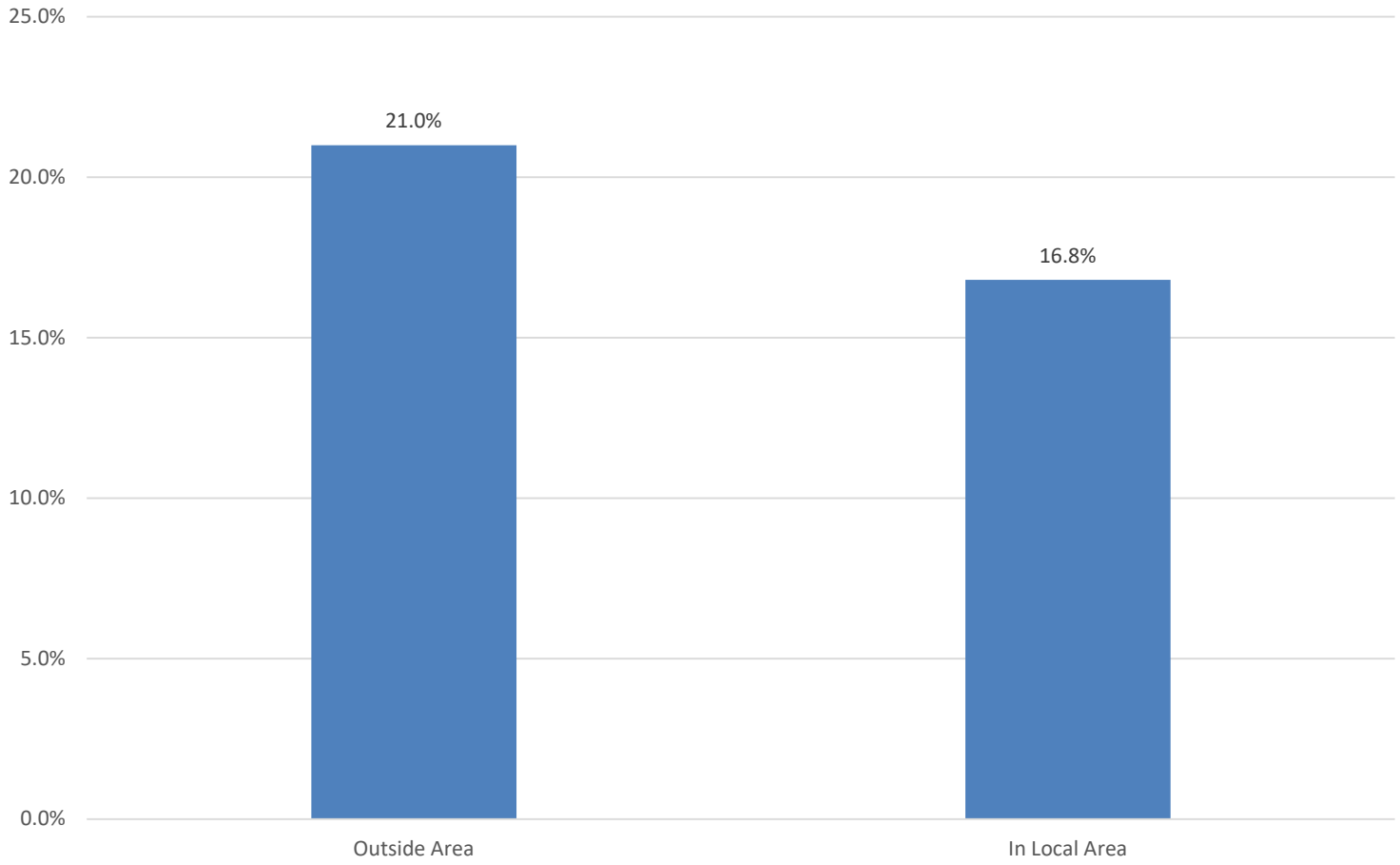
4-yr graduation rate by parental education



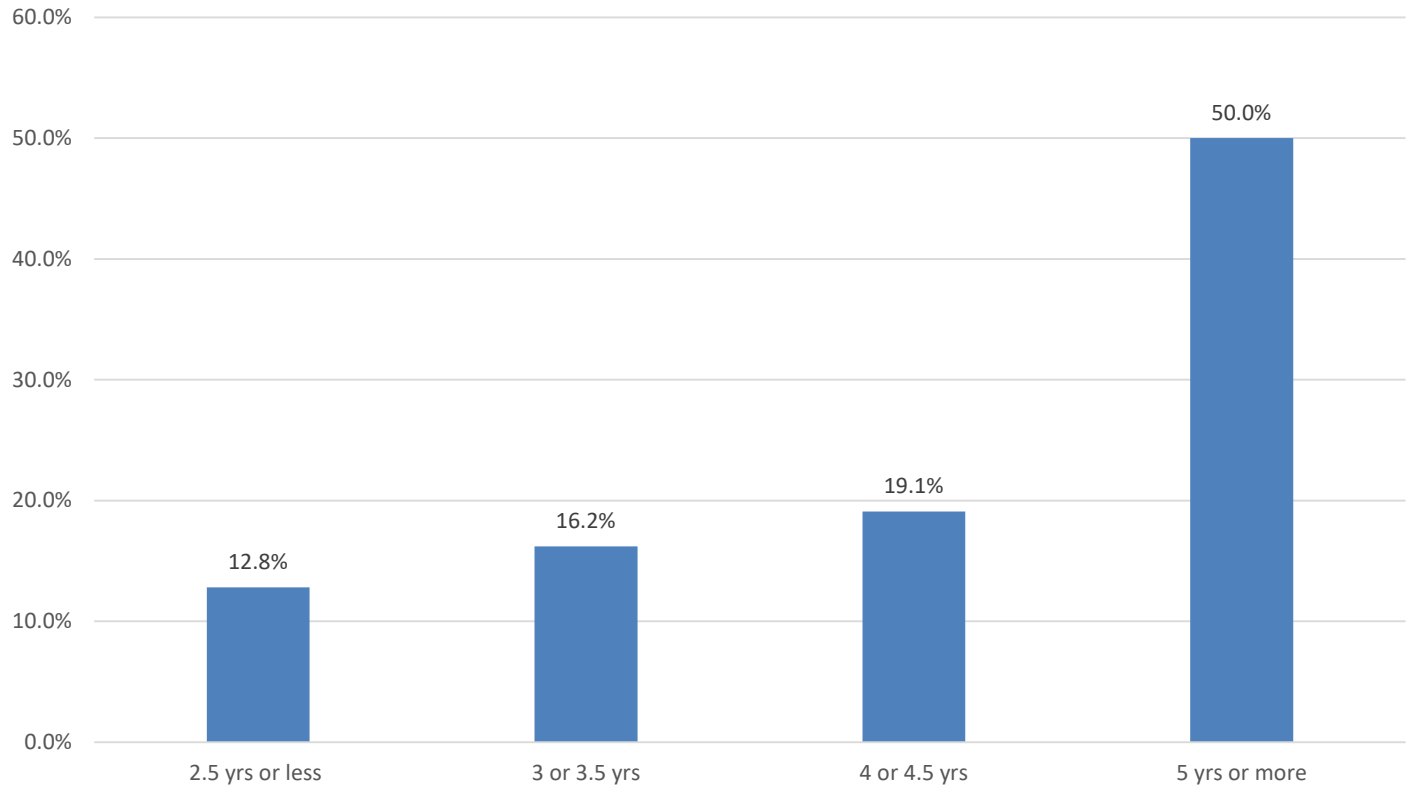
4-yr graduation rate by Pell Grant



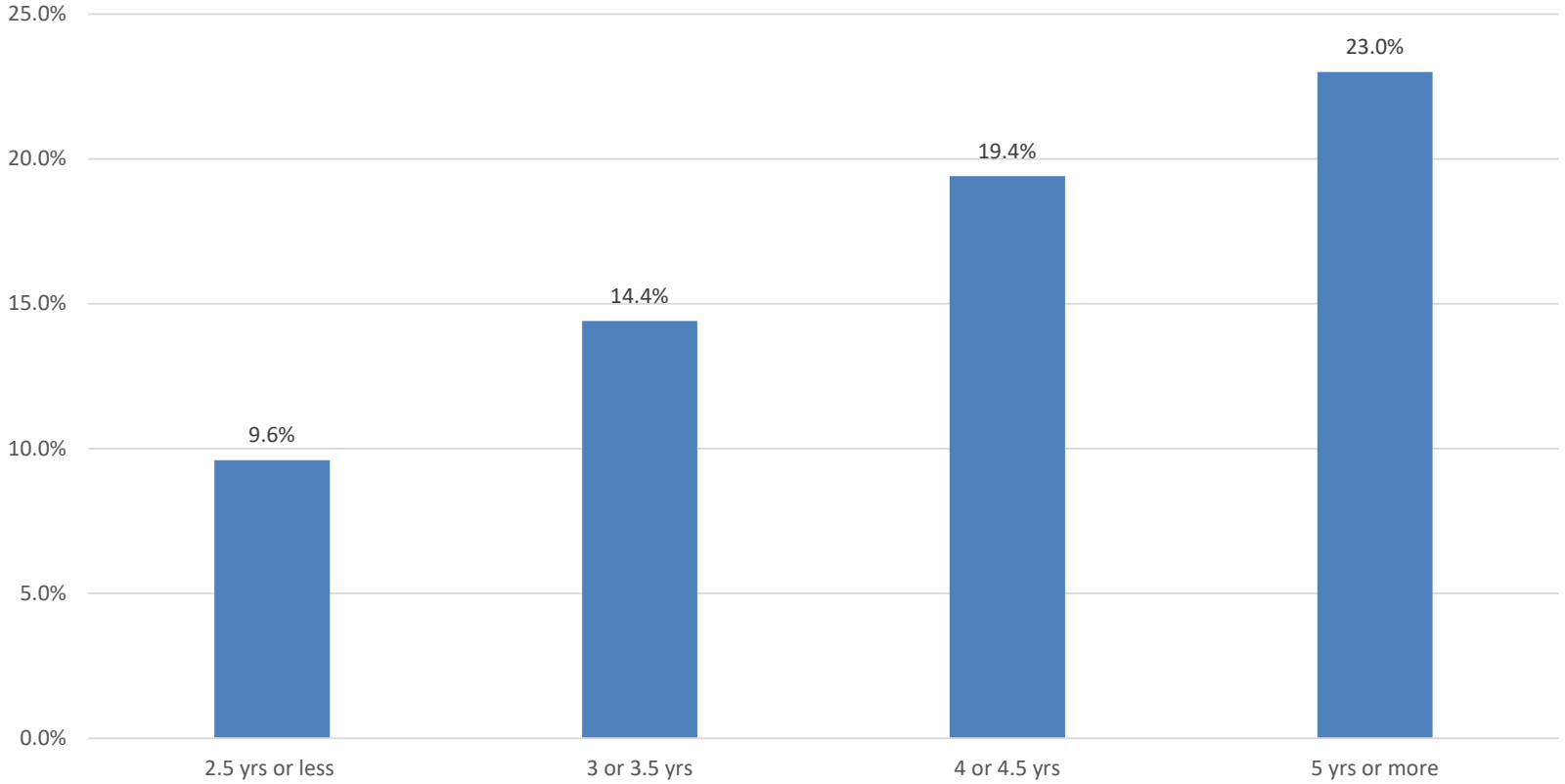
4-yr graduation rate by admission area



4-yr graduation rate by college preparatory math courses



4-yr graduation rate by college preparatory English courses



Method

- Logistic regression
  - Dependent variable:
    - Graduated in 4 years (yes/no)
  - Independent variables
    - Categorical:
      - Gender, ethnicity, parental education, major, admission area, Pell Grant received.
      - College preparatory courses
    - Continuous:
      - HS GPA, units taken during the first time, SAT (ACT converted if SAT missing)

Result summary

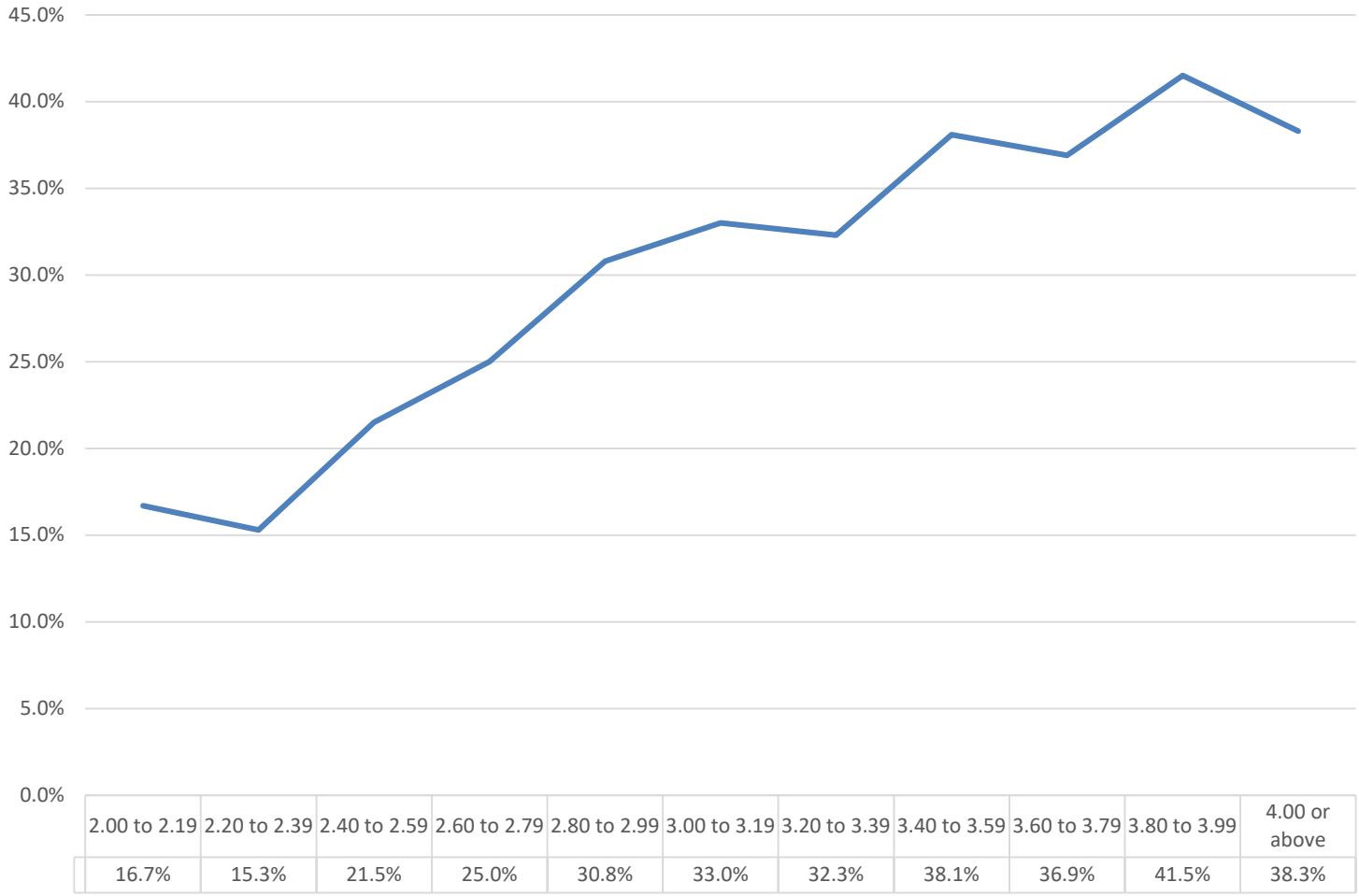
- Nagelkerke R^2: 21.5%
- 70% overall accuracy

	Predictor	Odds Ratio	Sig.
	HS GPA	1.4	<0.001
	Units taken, first term	1.4	<0.001
	SAT	1.2	<0.001
	Men	0.5	<0.001
	Local	0.9	<0.001
	Pell Received	0.7	<0.001
Ethnicity	American Ind.	0.7	0.281
	Black	0.8	0.028
	Hispanic	0.6	<0.001
	Asian	0.7	<0.001
	Non-Res	0.7	0.035
	Pacific Islander	0.3	0.032
	Unknown	0.7	<0.001
parent ed	First Generation	0.7	<0.001
	Some College	0.8	<0.001
	Unknown	0.9	0.149
Major	Arts	0.3	<0.001
	Business	0.7	<0.001
	Communications	1.3	0.001
	Engineering	0.1	<0.001
	Health and Human Dev	0.6	<0.001
	Natural Science/Math	0.3	<0.001
	Other	0.5	<0.001
College Preparatory	English	1.0	0.953
	Math	1.1	<0.001
	History	1.1	0.037
	Social Science	1.0	0.337
	Foreign Language	1.1	0.001
	Arts	1.0	0.722
	Electives	1.2	0.072

Two-year graduation rates by transfer student characteristics

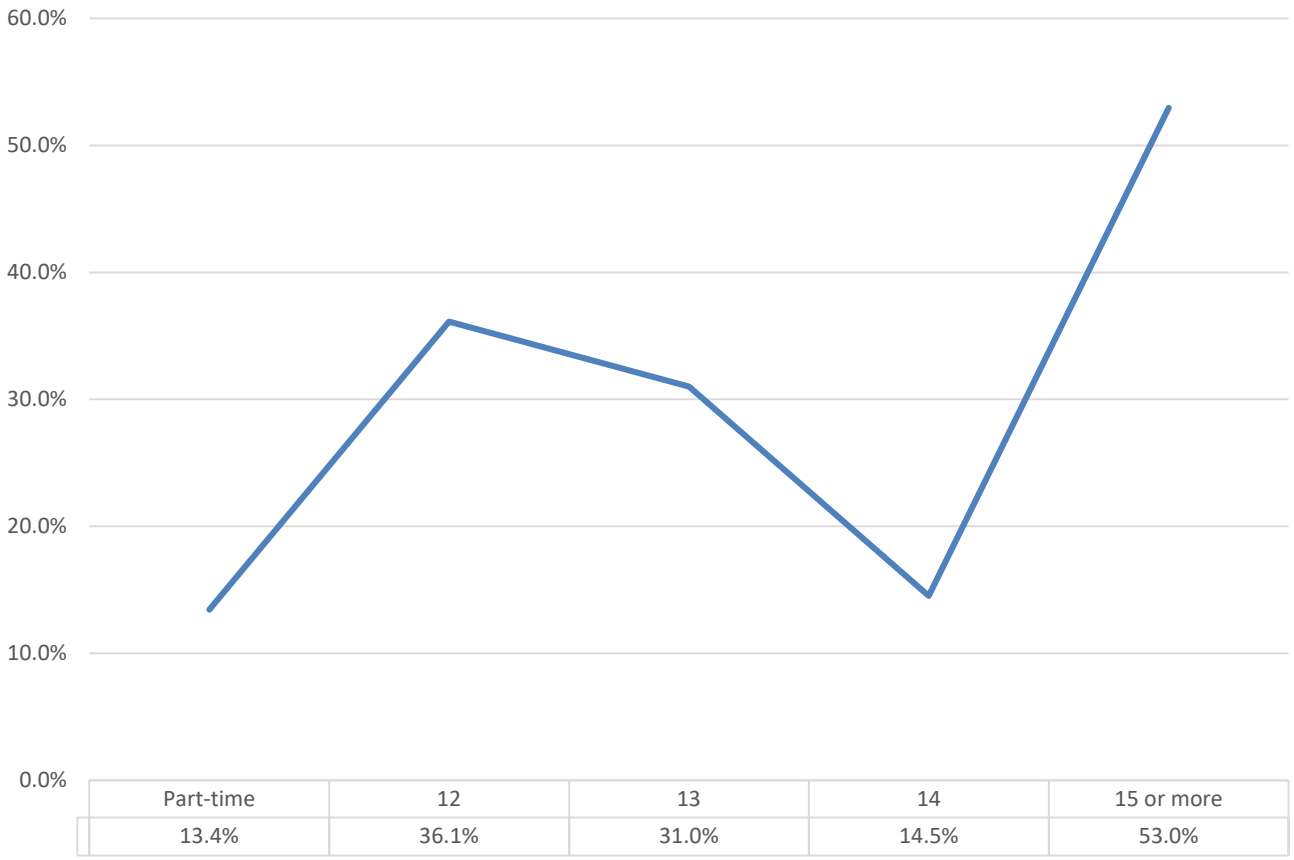
- Data
  - Transfer students from fall 2012 to fall 2014 cohorts (including spring)
    - 13848 Transfer students
    - 4492 (32.4%) graduated in 2 years or less

2-yr graduation rate by Transfer GPA



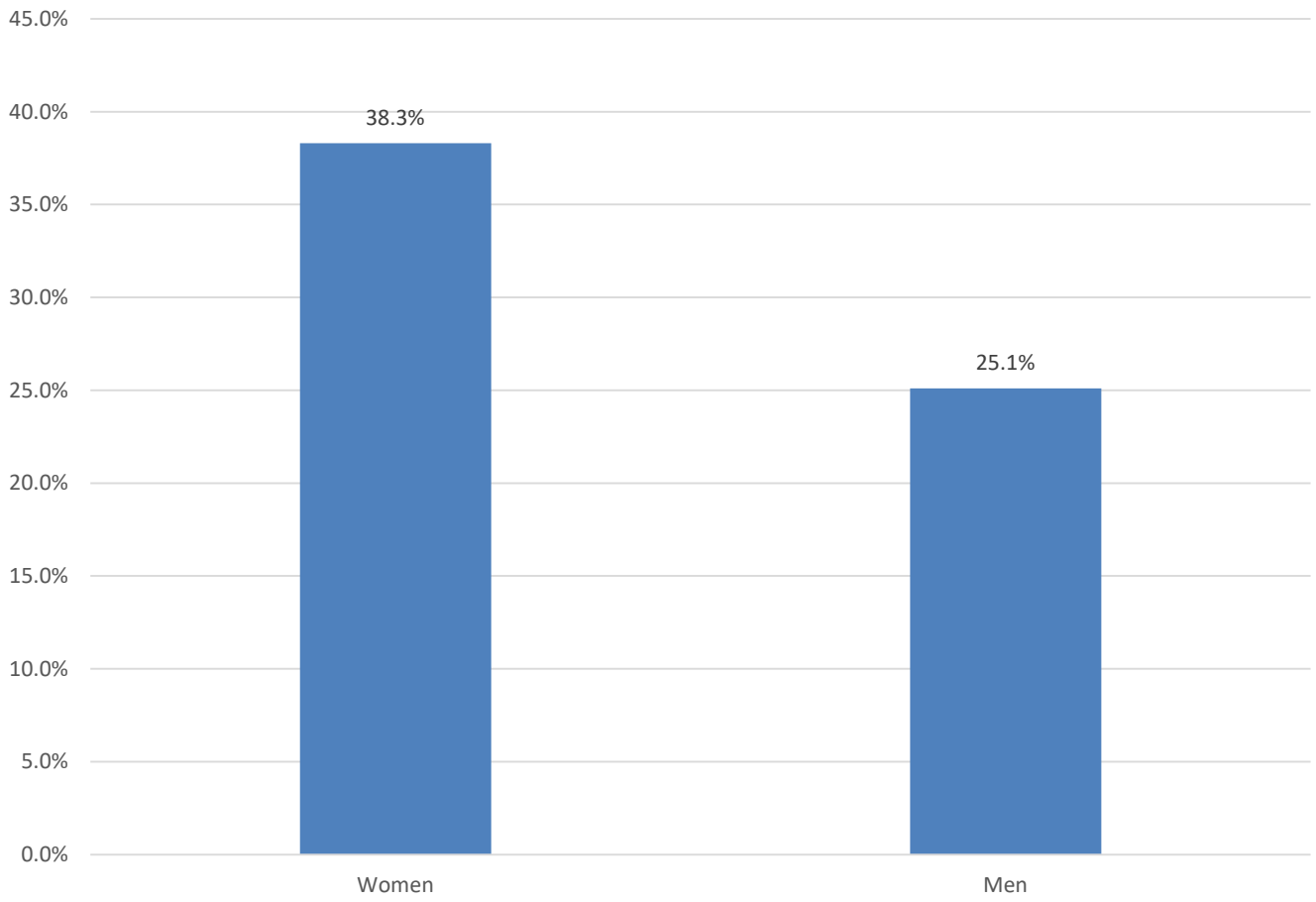
2.00 to 2.19	2.20 to 2.39	2.40 to 2.59	2.60 to 2.79	2.80 to 2.99	3.00 to 3.19	3.20 to 3.39	3.40 to 3.59	3.60 to 3.79	3.80 to 3.99	4.00 or above
16.7%	15.3%	21.5%	25.0%	30.8%	33.0%	32.3%	38.1%	36.9%	41.5%	38.3%

2-yr graduation rate by 1st Term Units

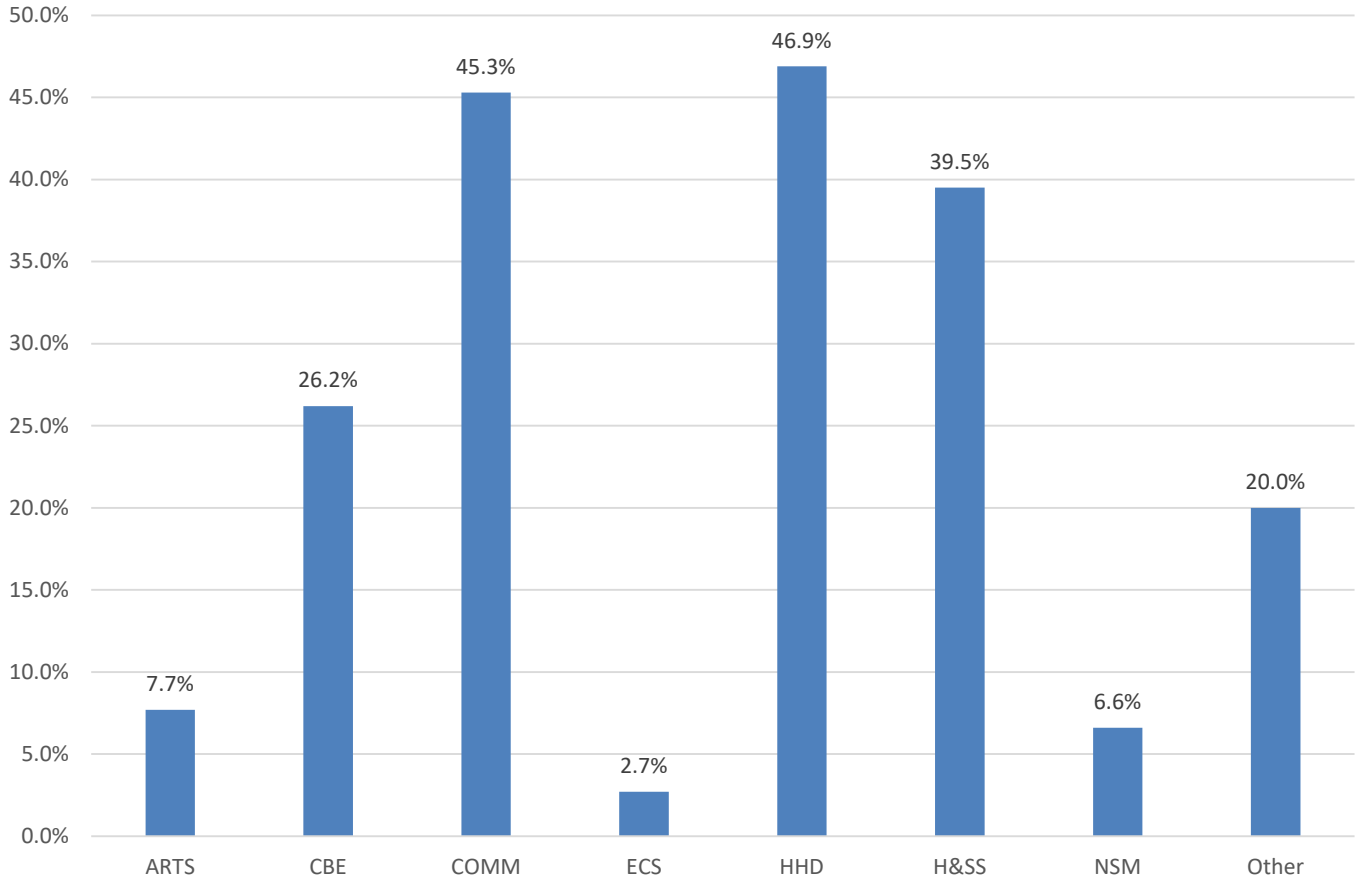


Part-time	12	13	14	15 or more
13.4%	36.1%	31.0%	14.5%	53.0%

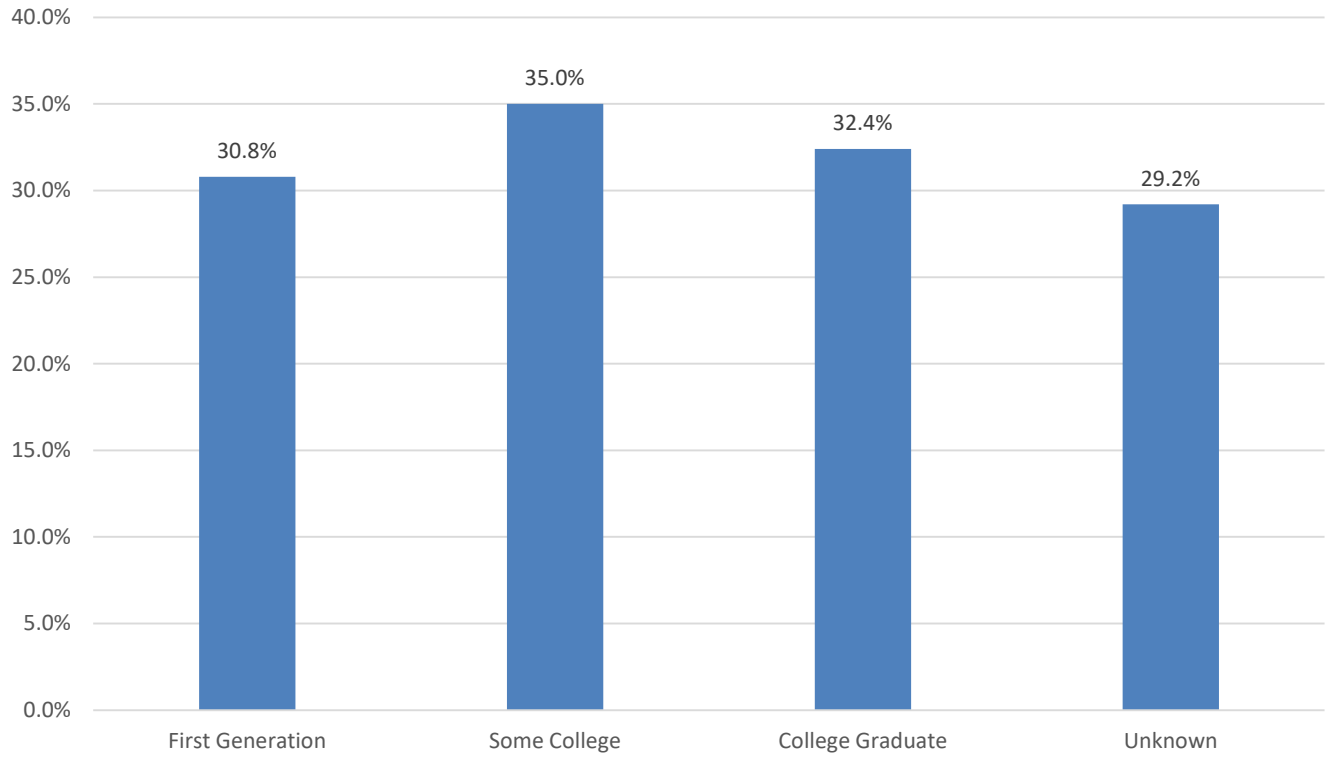
2-yr graduation rate by gender



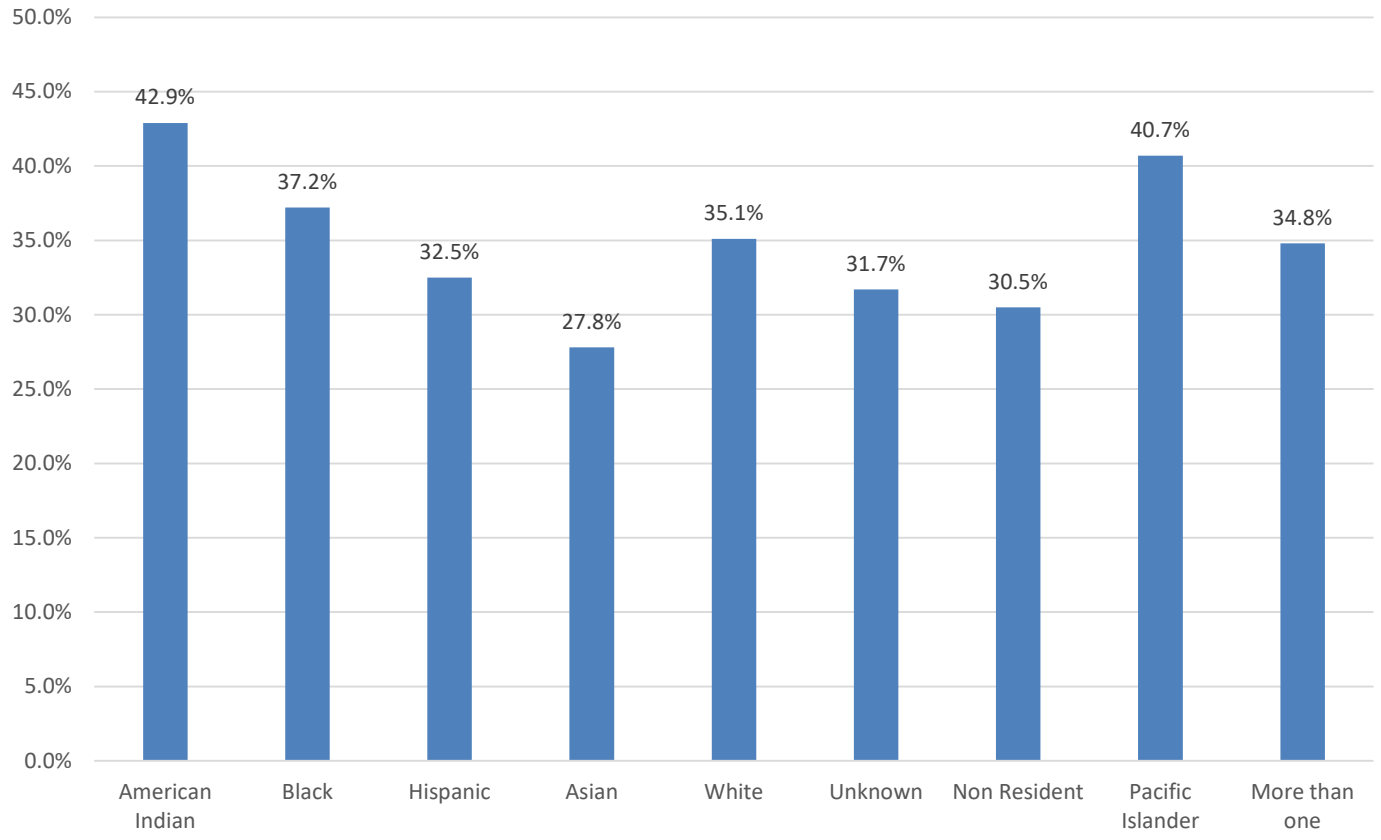
2-yr graduation rate by major at entry



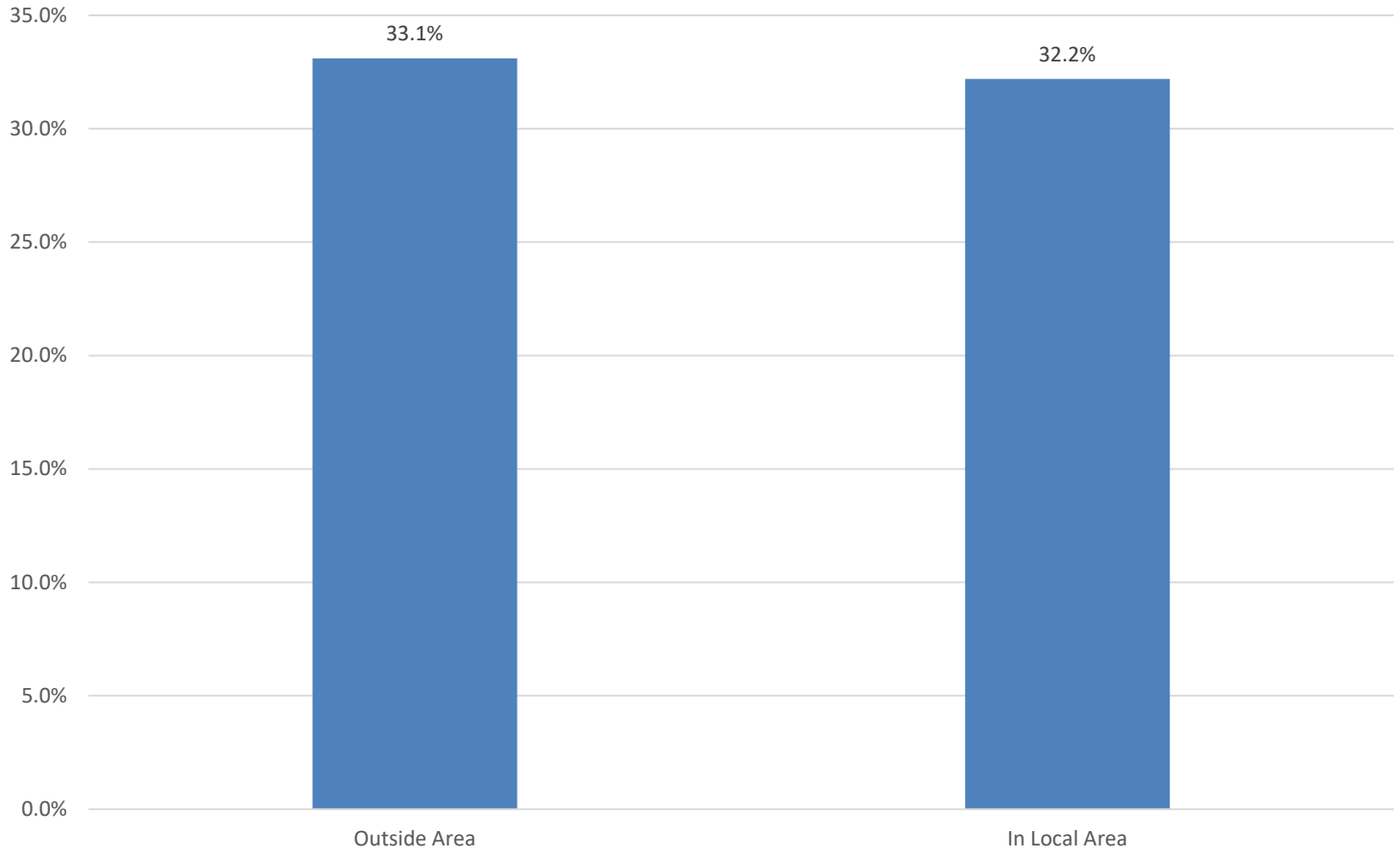
2-yr graduation rate by parental education



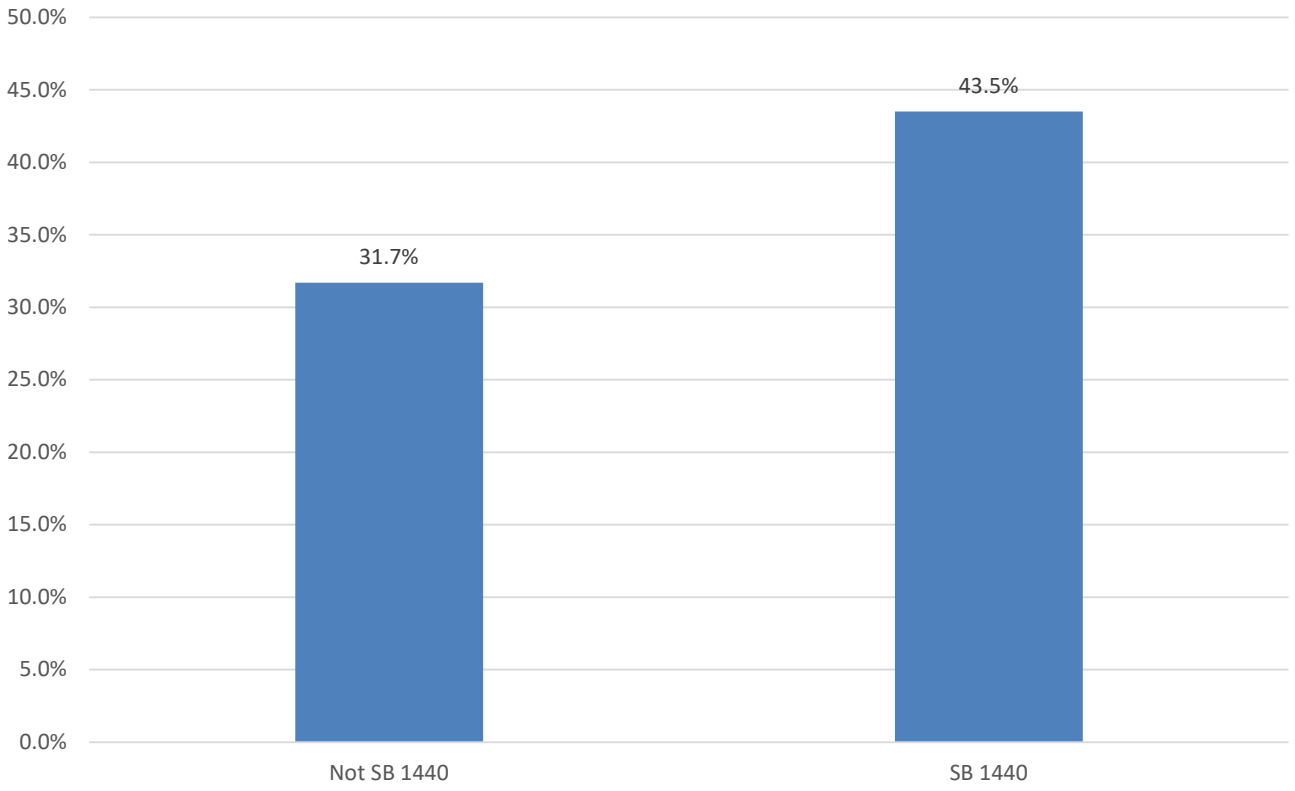
2-yr graduation rate by ethnicity



2-yr graduation rate by admission area



2-yr graduation rate by SB 1440 status



Method

- Logistic regression
  - Dependent variable:
    - Graduated in 2 years (yes/no)
  - Independent variables
    - Categorical:
      - Gender, ethnicity, parental education, major, admission area, Pell Grant received, Fall entry, SB 1440.
    - Continuous:
      - Transfer GPA, units taken during the first time, Transfer Units

Results summary

- Nagelkerke R^2: 29.3%
- 70% overall accuracy

	Predictor	Odds Ratio	Sig.
	Transfer GPA	1.3	<0.001
	Transfer Units	1.3	<0.001
	Units taken, first term	2.4	<0.001
	Fall entry	1.2	0.066
	Men	0.7	<0.001
	local	1.3	0.002
	SB1440 status	1.3	<0.001
	Pell Received	0.8	<0.001
Ethnicity	American Ind.	1.7	0.213
	Black	1.0	0.757
	Hispanic	0.9	0.104
	Asian	0.8	<0.001
	Non-Res	0.9	0.28
	Pacific Islander	1.4	0.43
	Unknown	0.9	0.476
parent ed	First Generation	1.0	0.665
	Some College	1.1	0.031
	Unknown	0.8	0.041
Major	Arts	0.1	<0.001
	Business	0.5	<0.001
	Communications	1.2	0.022
	Engineering	0.03	<0.001
	Health and Human Dev	1.3	<0.001
	Natural Science/Math	0.1	<0.001
	Other	0.3	0.299

Summary

- ✓ Less well academically prepared students are less likely to succeed.
- ✓ Men are less likely to succeed than women
- ✗ URM are less likely to graduate
- ✓ Higher # of Units in 1st term is associated with graduation
- ✓ STEM majors less likely to succeed
- ✗ Students from outside of our local admissions area are less likely to succeed
- △ Freshmen who complete remediation in mathematics early in their academic careers are more likely to succeed.

More options

- Feature engineering
- Discrete-time survival analysis
- Competing risk analysis
- Survival Tree

Feature Engineering

Feature engineering: create new variables from existing ones.

Example: application data

Quantify institutional characteristics

Institution	Mean SAT	Mean HS GPA	First Gen %	URM %	admission rate	application count
A	1050	3.23	0.18	0.1	0.6	73
B	1092.38	2.99	0.06	0.04	0.46	96
C	1065.73	3.25	0.06	0.06	0.67	78
D	1055.37	3.09	0.18	0.18	0.67	45
E	1068.68	2.98	0.02	0.05	0.43	42

- These predictors were added to the FTF model for 4-year graduation
  - Nagelkerke R^2: 23%
  - 71% overall accuracy

Feature Engineering

Feature engineering: create new variables from existing ones.

Example: DFW

1. Calculate average DFW for courses taken in the first term
2. Calculate Max values of DFW among the courses taken
3. Count the number of courses with DFW >=20%

student	Course	DFW
1	DANC 101	3.0%
1	HCOM 100	10.0%
1	HIST 110B	14.0%
1	MATH 115	30.0%
2	ENGL 99	12.0%
2	HIST 110B	14.0%
2	MATH 115	30.0%
2	POSC 100	22.0%

student	Avarage DFW	Max DFW	Count > 20 %
1	14.3%	30.0%	1
2	19.5%	30.0%	2

Feature Engineering

Feature engineering: create new variables from existing ones.

Example: DFW.

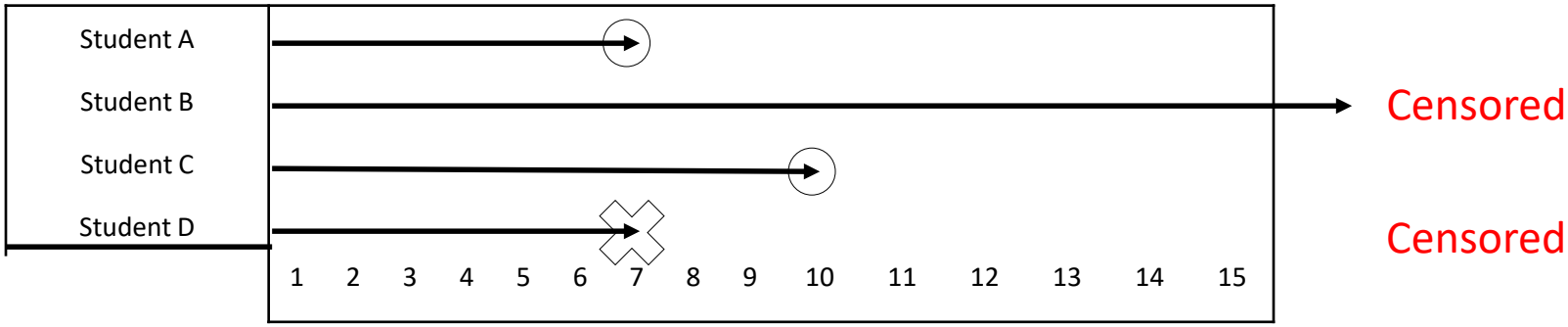
1. Calculate average DFW for courses taken in the first term
2. Calculate Max values of DFW among the courses taken
3. Count the number of courses with DFW >=20%

- These predictors were added to the FTF model for 4-year graduation
  - Nagelkerke R^2: 22%
  - 71% overall accuracy

Predictor	Odds Ratio	Sig.
Average DFW by faculty	1.0	0.5
Average DFW by course	0.9	<0.001
Max DFW by faculty	0.9	0.18
Max DFW by course	0.9	0.01
Count of high DFW by faculty	1.0	0.35
Count of high DFW by course	1.3	<0.001

Survival analysis

- This can be used to examine the length of time until the occurrence of some event, such as time to degree.
- Two possible outcomes:
  - event happened
  - Not happened (censored)



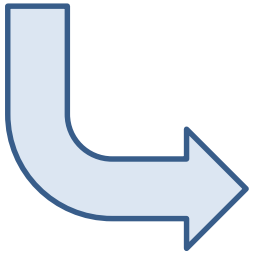
Graduated

Dropped out

	Graduated	Term
Student A	1	7
Student B	0	15
Student C	1	10
Student D	0	7

Data structure

Student ID	last_term	Graduated	Male
1	4	0	0
2	2	0	0
3	7	1	1



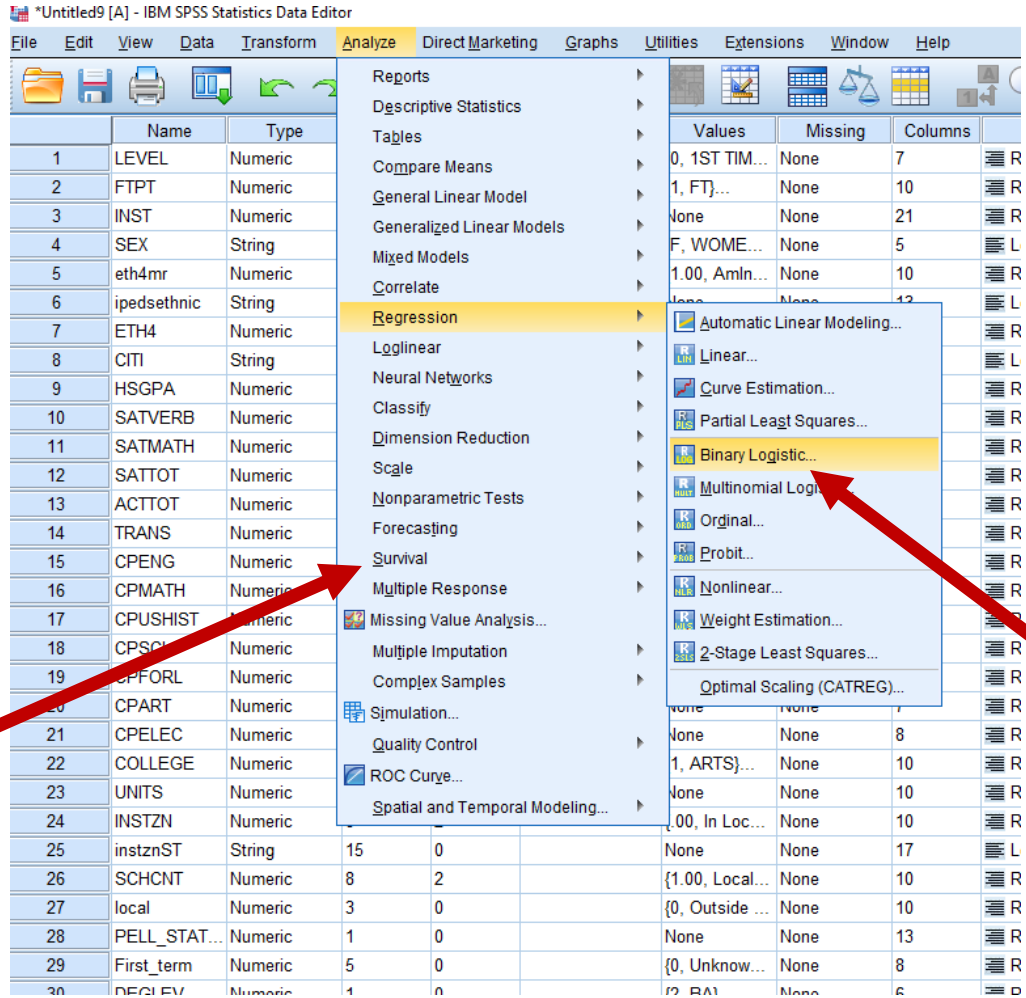
DV

Binary variables for each term

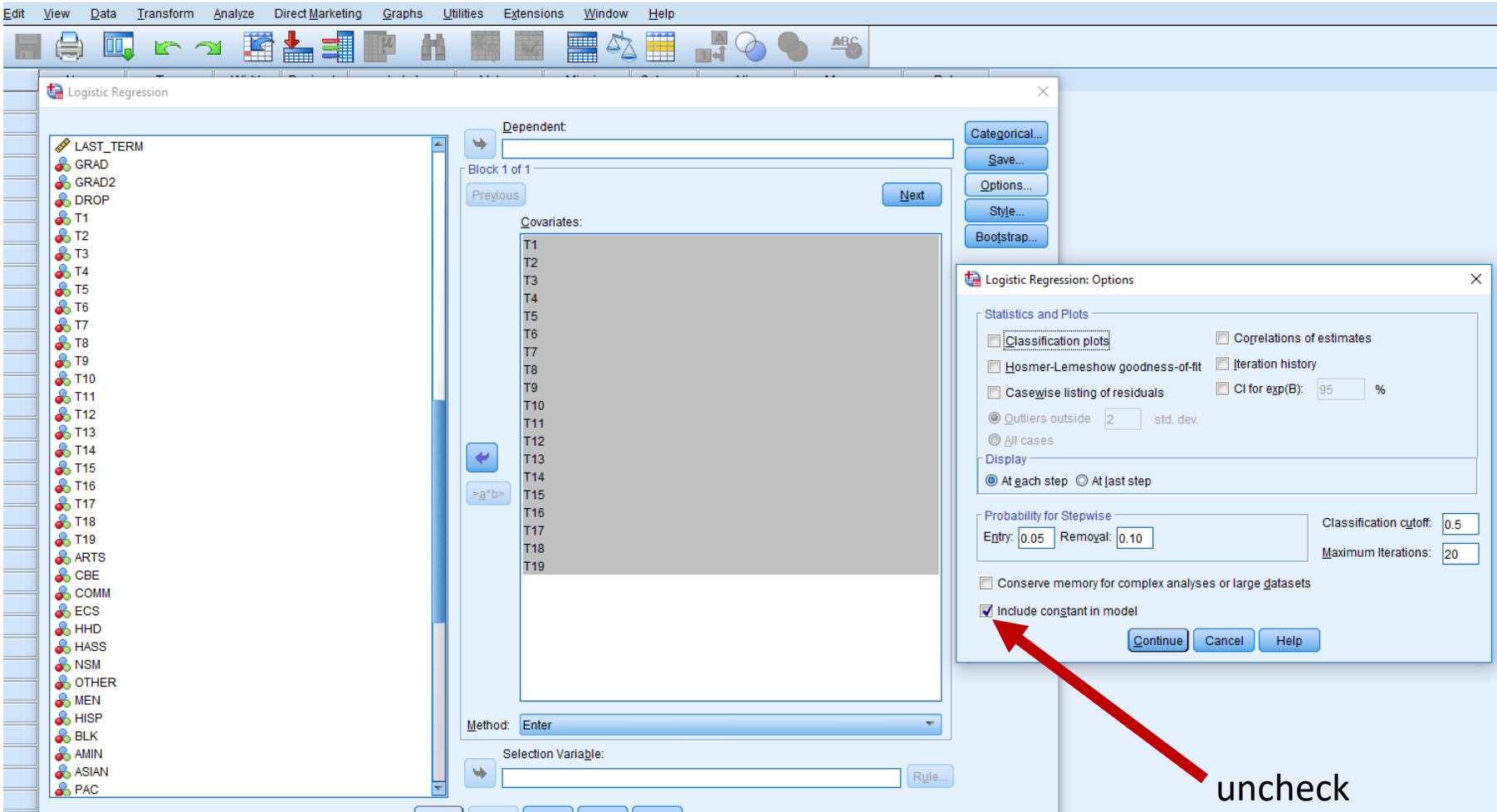
Student ID	Term	Graduated	Male	T1	T2	T3	T4	T5	T6
1	1	0	0	1	0	0	0	0	0
1	2	0	0	0	1	0	0	0	0
1	3	0	0	0	0	1	0	0	0
1	4	0	0	0	0	0	1	0	0
2	1	0	1	1	0	0	0	0	0
2	2	0	1	0	1	0	0	0	0
3	1	0	0	1	0	0	0	0	0
3	2	0	0	0	1	0	0	0	0
3	3	0	0	0	0	1	0	0	0
3	4	0	0	0	0	0	1	0	0
3	5	0	0	0	0	0	0	1	0
3	6	0	0	0	0	0	0	0	1
3	7	1	0	0	0	0	0	0	0

In spss

Not this,
Unless time is
continuous



this



In R, include **-1** in the formula. For eample, glm(GRAD ~ T1 + T2 + T3… - 1, family="binomial")

Survival analysis

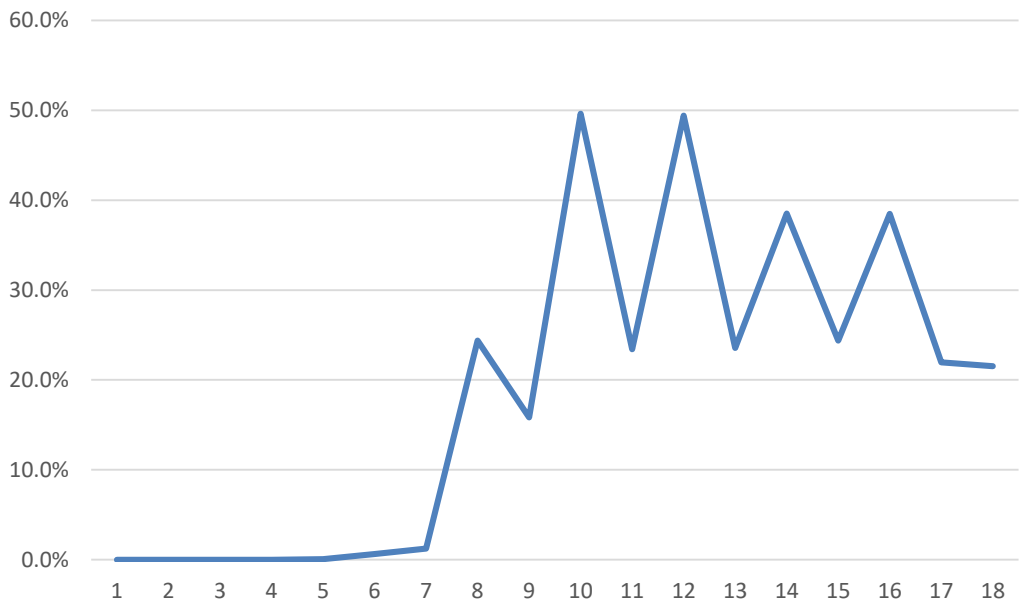
- Data
  - FTF from fall 2008 to fall 2013 cohorts
    - 26059 FTFs.
- Method
  - Discrete-time survival analysis
    - DV: Graduated = 1; Not graduated = 0
    - IV: Use the same set of predictors as logistic regression analysis plus 18 binary time indicator variables.
      - Step 1: time indicators only
      - Step 2: enter all the other predictors.

Survival analysis

Step 1: Time Indicator variables only

Term	B	hazard
T1	-21.2	0.0%
T2	-21.2	0.0%
T3	-10.0	0.0%
T4	-10.0	0.0%
T5	-7.5	0.1%
T6	-5.1	0.6%
T7	-4.4	1.2%
T8	-1.1	24.4%
T9	-1.7	15.8%
T10	0.0	49.6%
T11	-1.2	23.4%
T12	0.0	49.4%
T13	-1.2	23.6%
T14	-0.5	38.5%
T15	-1.1	24.4%
T16	-0.5	38.5%
T17	-1.3	21.9%
T18	-1.3	21.5%

Hazard is the probability of an event during the term given the event hasn't occured earlier. Convert B by 1/(1 + EXP(B))



Survival analysis

Step 2: Add predictors.

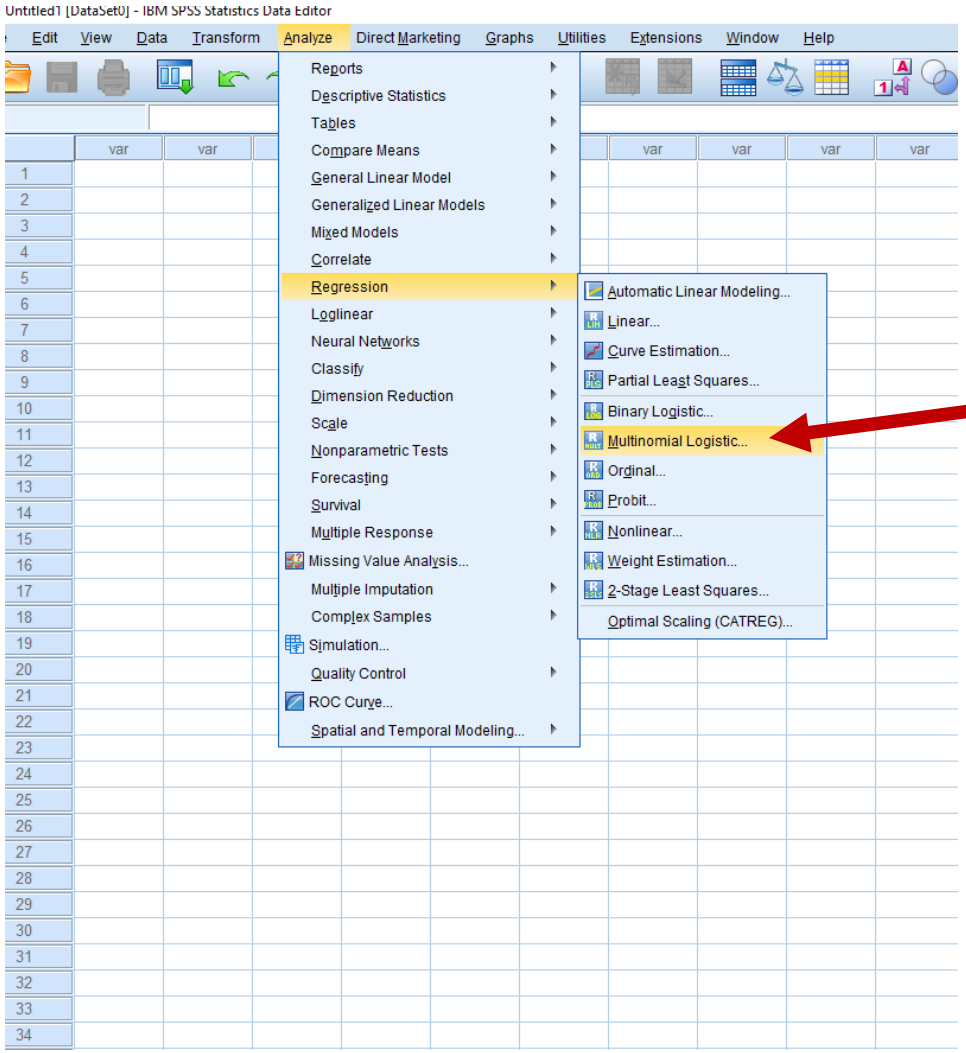
	Predictor	Odds Ratio	Sig.
	HS GPA	1.189	<0.001
	Units taken	1.224	<0.001
	SAT	1.071	<0.001
	Men	0.69	<0.001
	Local	0.874	<0.001
	Pell Received	0.856	<0.001
Ethnicity	American Ind.	1.028	0.902
	Black	0.739	<0.001
	Hispanic	0.663	<0.001
	Asian	0.736	<0.001
	Non-Res	0.662	<0.001
	Pacific Islander	0.643	0.06
	Unknown	0.827	<0.001
parent ed	First Generation	0.799	<0.001
	Some College	0.866	<0.001
	Unknown	0.907	0.046
Major	Arts	0.518	<0.001
	Business	0.773	<0.001
	Communications	1.289	<0.001
	Engineering	0.366	<0.001
	Health and Human Dev	0.794	<0.001
	Natural Science/Math	0.413	<0.001
	Other	0.65	<0.001
College Preparatory	English	0.917	0.006
	Math	1.043	0.011
	History	1.054	0.038
	Social Science	0.984	0.345
	Foreign Language	1.024	0.122
	Arts	0.979	0.197
	Electives	0.955	0.355

More options

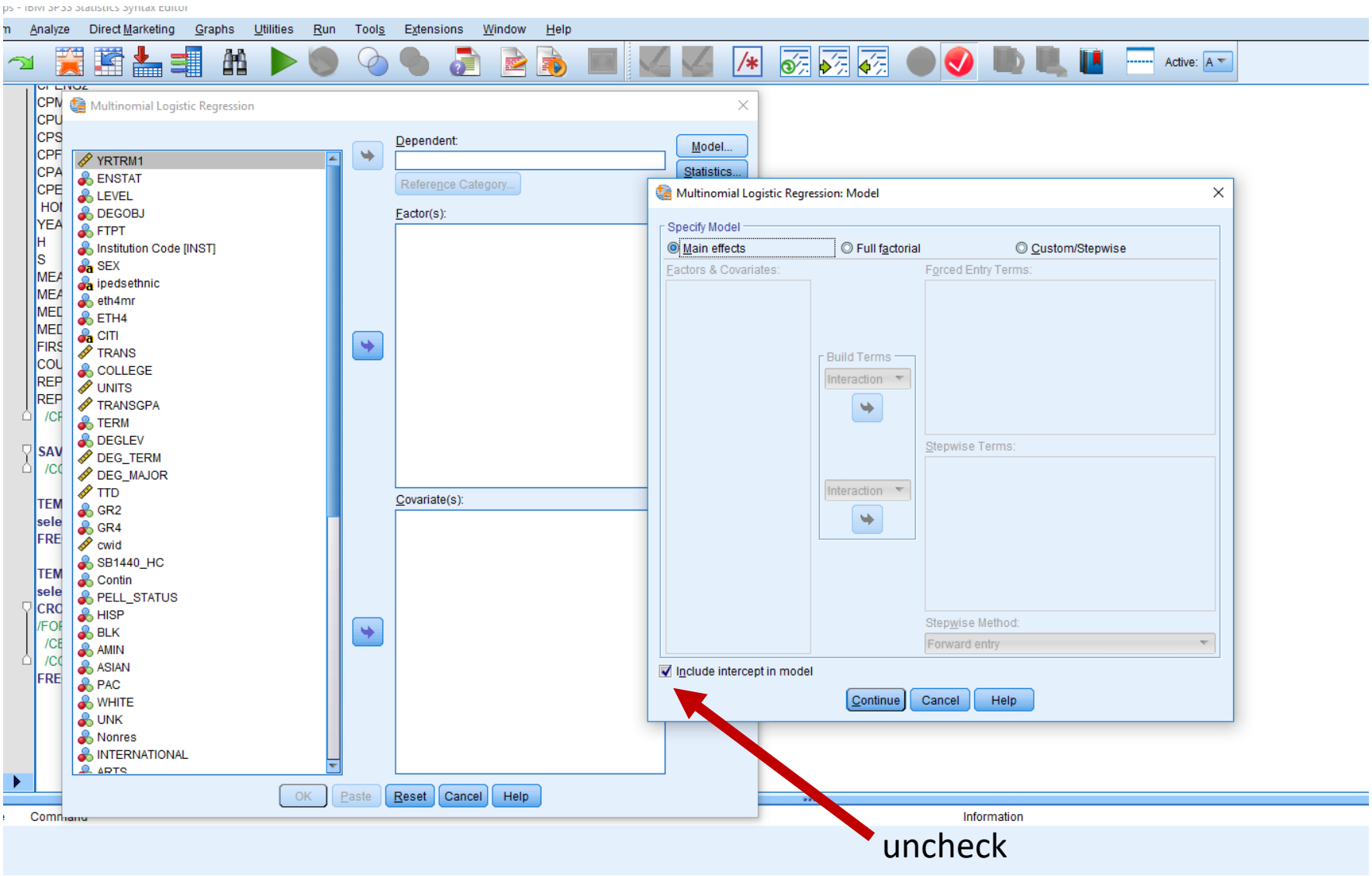
- Competing risk analysis
  - Use survival data
  - More than two outcomes
    - Graduated = 1, dropped out = 2, neither = 0.

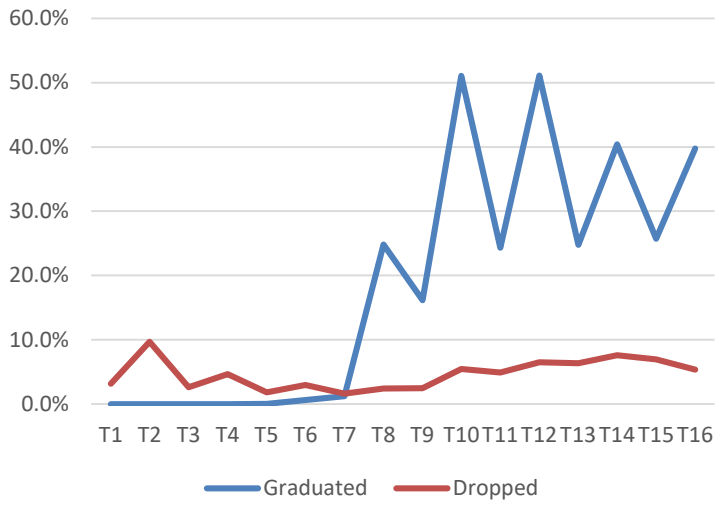
- Machine learning
  - Survival data can be examined more modern methods.
  - In R, for example, you can use *DStree* to build *discrete-time survival tree*.

In SPSS


Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor
Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Optimal Scaling (CATREG)...
var

this

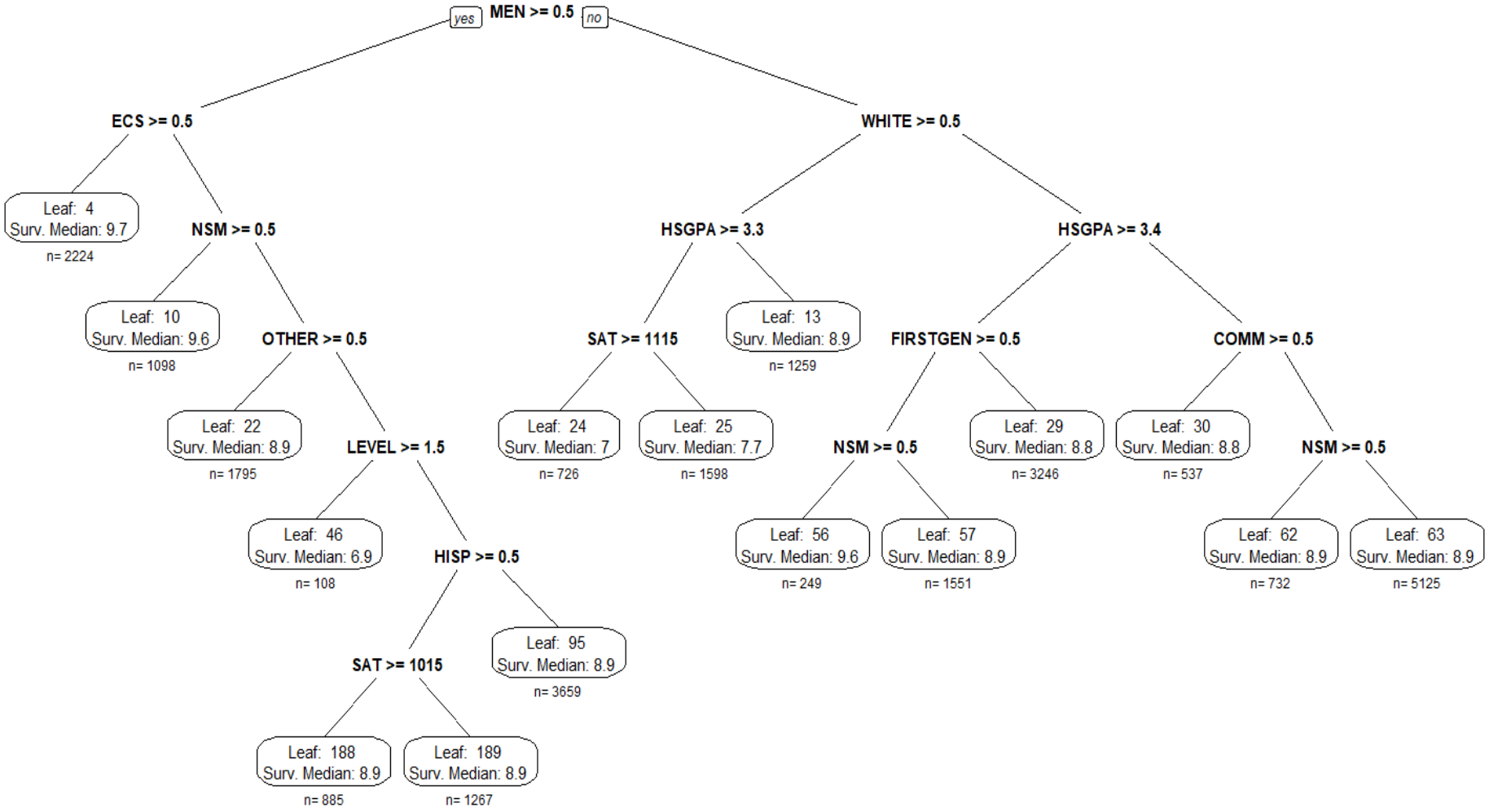

Analyze
Direct Marketing
Graphs
Utilities
Run
Tools
Extensions
Window
Help
Multinomial Logistic Regression
YRTRM1
ENSTAT
LEVEL
DEGOBJ
FTPT
Institution Code [INST]
SEX
ipedsethnic
eth4mr
ETH4
CITI
TRANS
COLLEGE
UNITS
TRANSGPA
TERM
DEGLEV
DEG_TERM
DEG_MAJOR
TTD
GR2
GR4
cwid
SB1440_HC
Contin
PELL_STATUS
HISP
BLK
AMIN
ASIAN
PAC
WHITE
UNK
Nonres
INTERNATIONAL
Dependent:
Reference Category...
Factor(s):
Covariate(s):
Model...
Statistics...
OK
Paste
Reset
Cancel
Help
Multinomial Logistic Regression: Model
Specify Model
Main effects
Full factorial
Custom/Stepwise
Factors & Covariates:
Build Terms
Interaction
Forced Entry Terms:
Stepwise Terms:
Stepwise Method:
Forward entry
Include intercept in model
Continue
Cancel
Help
Information
uncheck



	6yr-grad rate
Black	47.00%
Hispanic	54.90%
Asian	64.50%
White	64.60%
Non Local	59.00%
Local	60.20%

		Graduated		Dropped out	
	Predictor	Odds Ratio	Sig.	Odds Ratio	Sig.
	HS GPA	1.2	<0.001	0.8	<0.001
	Units taken	1.2	<0.001	1.0	0.848
	SAT	1.1	<0.001	1.0	0.172
	Men	0.7	<0.001	1.2	<0.001
	Local	0.9	<0.001	0.7	<0.001
	Pell Received	0.9	<0.001	1.0	0.665
Ethnicity	American Ind.	1.0	0.855	1.3	0.174
	Black	0.7	<0.001	1.2	0.002
	Hispanic	0.7	<0.001	1.0	0.5
	Asian	0.7	<0.001	0.8	<0.001
	Non-Res	0.7	<0.001	1.2	0.083
	Pacific Islander	0.6	0.054	0.7	0.256
	Unknown	0.8	<0.001	1.1	0.089
parent ed	First Generation	0.8	<0.001	1.2	<0.001
	Some College	0.9	<0.001	1.1	<0.001
	Unknown	0.9	0.055	1.1	0.13
Major	Arts	0.5	<0.001	1.3	<0.001
	Business	0.8	<0.001	0.9	0.164
	Communications	1.3	<0.001	0.9	0.123
	Engineering	0.4	<0.001	1.2	<0.001
	Health and Human Dev	0.8	<0.001	1.0	0.466
	Natural Science/Math	0.4	<0.001	1.1	0.011
	Other	0.6	<0.001	1.0	0.448
College Preparatory	English	0.9	0.008	1.1	0.032
	Math	1.0	0.025	0.9	<0.001
	History	1.1	0.027	1.1	0.005
	Social Science	1.0	0.248	0.9	<0.001
	Foreign Language	1.0	0.163	0.9	0.001
	Arts	1.0	0.251	1.1	0.003
	Electives	1.0	0.351	1.0	0.836

Result of Discrete-time survival tree



End

Question?