# A MULTIVARIATE STATISTICAL ANALYSIS OF MAJOR CHANGE PATTERNS AND SIGNIFICANT FACTORS THAT INFLUENCE GRADUATION RATES: A CASE STUDY AT CALIFORNIA STATE UNIVERSITY, LONG BEACH

Yale Quan

Yale.Quan@csulb.edu

November 17, 2020

CALIFORNIA STATE UNIVERSITY
LONG BEACH

Chair: Dr. Sung Kim
Committee: Dr. Jen-Mei Chang, Dr. Kagba Suaray

# Definitions

- *First-time freshmen* – students who have never enrolled at another higher education institution before attending CSULB.

- *Cohort* - the group of students who started CSULB in the same academic year.

- *Time to Graduation* - the number of semesters it takes for a student to exit from CSULB:
    - Students who do not graduate
    - Students who graduated within four years (8 semesters)
    - Students whose time to graduation is longer than four years and no more than six years (12 semesters)

- *Graduation Rate* - The percentage of students who completed their degree within a specified time limit.

# **Background Information**

- In 2015 the California State University (CSU) system launched "Graduation Initiative 2025" which set a goal of
  - 40% four-year graduation rate
  - 70% six-year graduation rate
- There is a growing concern in the higher education community about how changing majors may affect timely graduation (Foraker, 2012; United States Department of Education 2017; Micceri 2011).
- There are also socioeconomic, demographic, and academic factors which might influence a student's time to graduation.

# The Dataset

- Obtained from CSULB Institutional Research and Analytics (IR&A) after obtaining approval from the Institutional Research Board.
- First-time freshman admitted between 2009 and 2012
  - Originally contained data on 16,468 first-time freshmen
  - 757,730 rows of course information
  - 181 variables containing academic and demographic information
- After determining the scope of the study I excluded students who:
  - are currently enrolled or enrolled for longer than 12 semesters
  - did not have a high school GPA and/or a standardized test score
  - were admitted into the university as a non-degree seeking student
  - enrolled in at least one Winter or Summer semester course

# Resulting Sample Sizes

| Cohort Year | Sample Size |
|:---:|:---:|
| 2009 | 1,792 |
| 2010 | 2,107 |
| 2011 | 2,655 |
| 2012 | 1,951 |
| **Total** | **8,505** |

# Research Question

- Does major change have a significant effect on a student's graduation status, and can this result be combined with the student's academic and demographic information to identify significant factors that influence timely graduation?

## Descriptive Statistics for Academic Variables

Measures of Academic Preparation:

- High School GPA
- SAT or ACT Reading Scores
- SAT or ACT Math Scores
- Eligibility Index Score

CSULB Academic Information:

- CSULB GPA
- DFW, WU, and WE courses
  - W = Withdrawn
  - WU = Unauthorized Withdrawal
  - WE = Extenuating Circumstances Withdrawal

**Goal:** To determine if there is a statistically significant difference between the cohorts

# Descriptive Statistics for Academic Variables

The distribution of High School GPA, SAT Reading Scores, SAT Math Scores, the number of DFW , WE, and WU courses, and overall CSULB GPA were reasonably homogenous between cohort years.

Therefore, I concluded that there was no statistically significant difference between the cohort years for these variables

## **Eligibility Index**

Eligibility Index (EI) is used to establish baseline admission eligibility for applicants to CSULB.

General EI $= 800$(High School GPA) $+$ SAT Critical Reading $+$ SAT Math

STEM EI $= 600$(High School GPA) $+$ SAT Critical Reading $+ 2$(SAT Math)

Note: SAT and ACT scores were converted to corresponding 2016 SAT scores

# Distribution of Eligibility Index Scores



| Cohort Year | Mean Eligibility Index Score | Median Eligibility Index Score | Standard Deviation |
|---|---|---|---|
| 2009 | 3852.39 | 3850 | 386.49 |
| 2010 | 3842.24 | 3862 | 362.69 |
| 2011 | 3818.56 | 3792 | 361.45 |
| 2012 | 3904.83 | 3912 | 359.36 |

## **Graduation and Major Change Rates**

- I visually explored the graduation and major change rates within each cohort year.

- The goal of this section is to visually determine if there appears to be any significance difference between the cohorts.  I also wanted to visually explore the relationship between major change and time to graduation

## Descriptive Statistics: Major Change Groupings

- After performing a review of the current literature, 4 major change groups were created:
    - A. Did Not Change Majors
    - B. Changed Majors Once
    - C. Changed Majors Twice
    - D. Changed Majors more than Twice
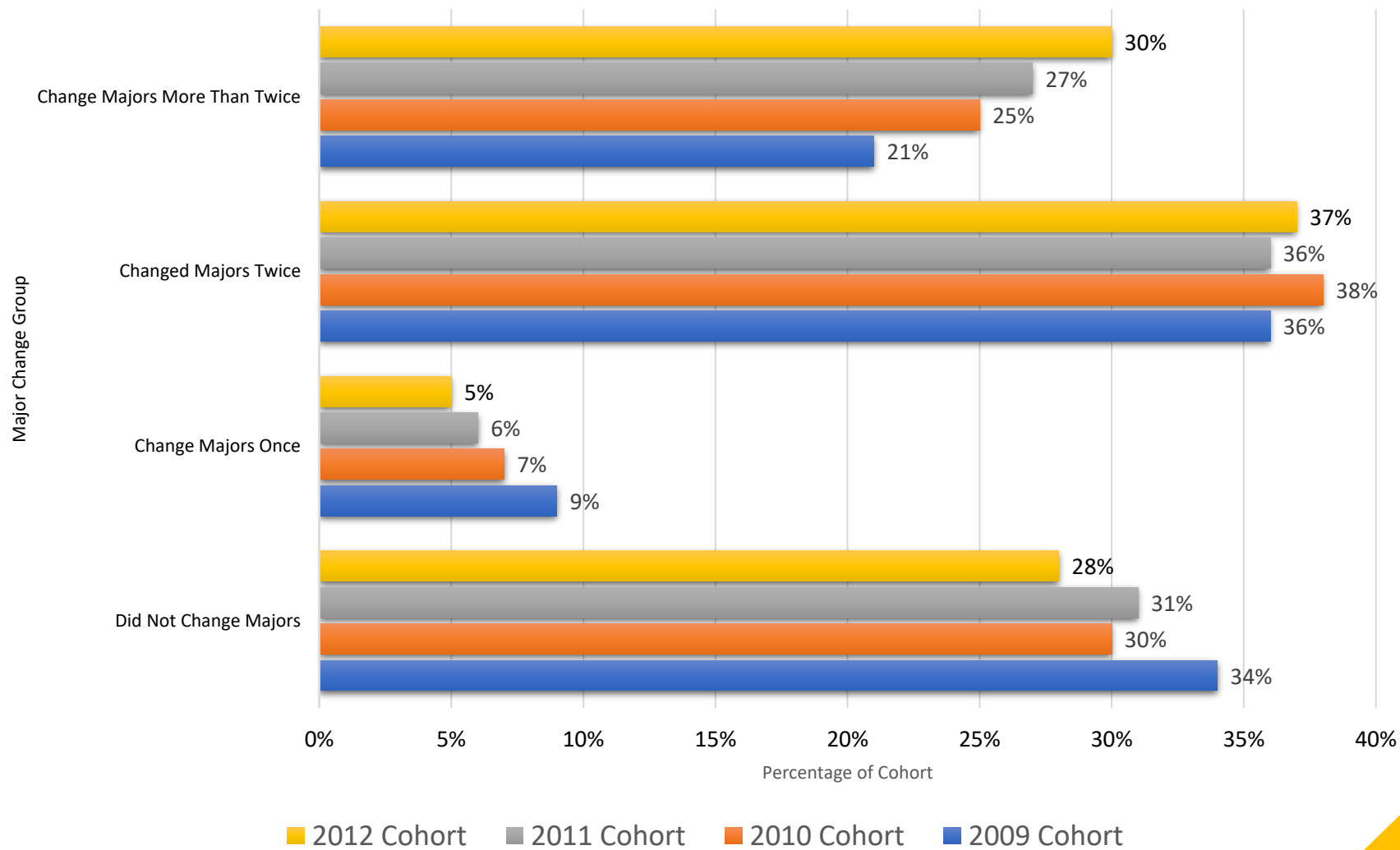
# Graduation Rates Per Cohort
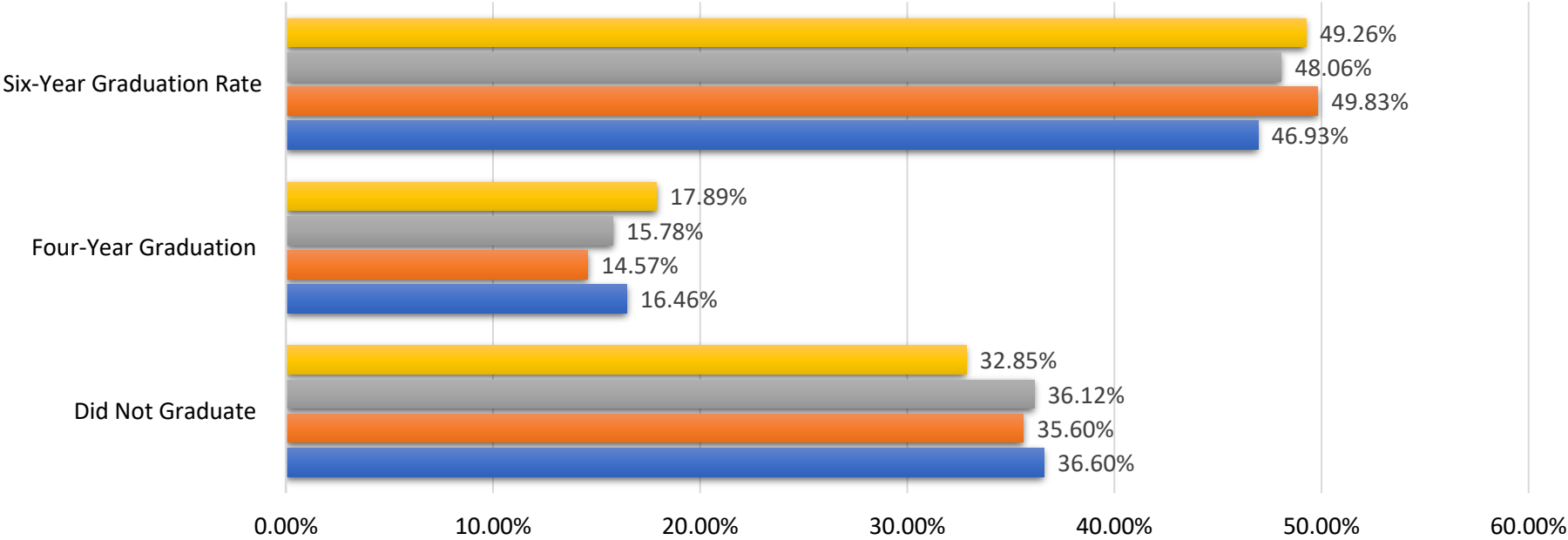


| | Did Not Graduate | Four-Year Graduation | Six-Year Graduation Rate |
|---|---|---|---|
| 2012 Cohort | 32.85% | 17.89% | 49.26% |
| 2011 Cohort | 36.12% | 15.78% | 48.06% |
| 2010 Cohort | 35.60% | 14.57% | 49.83% |
| 2009 Cohort | 36.60% | 16.46% | 46.93% |

Percentage of Cohort

■ 2012 Cohort  ■ 2011 Cohort  ■ 2010 Cohort  ■ 2009 Cohort
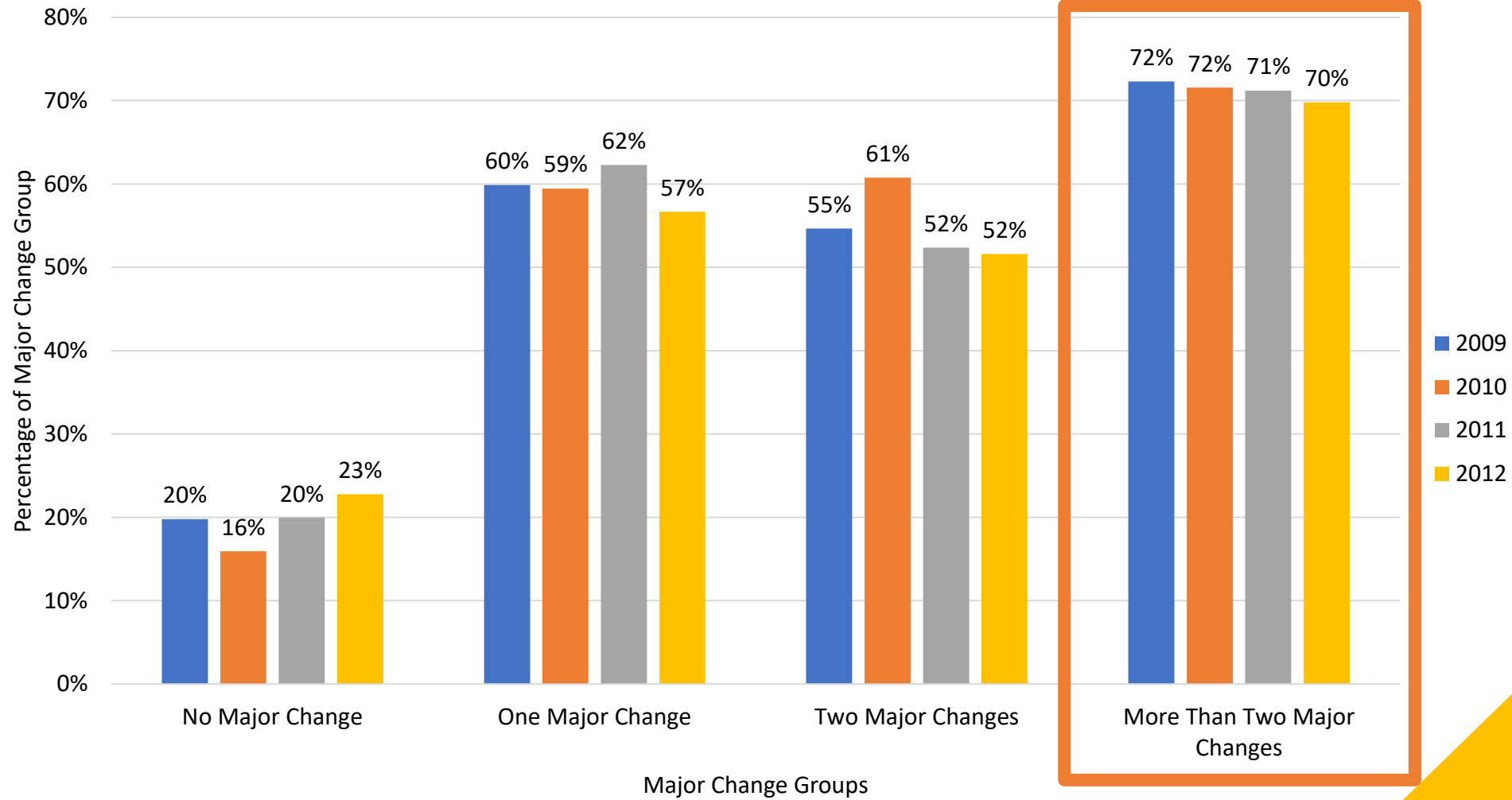
Four Year Graduation Rate Within Major Change Groups
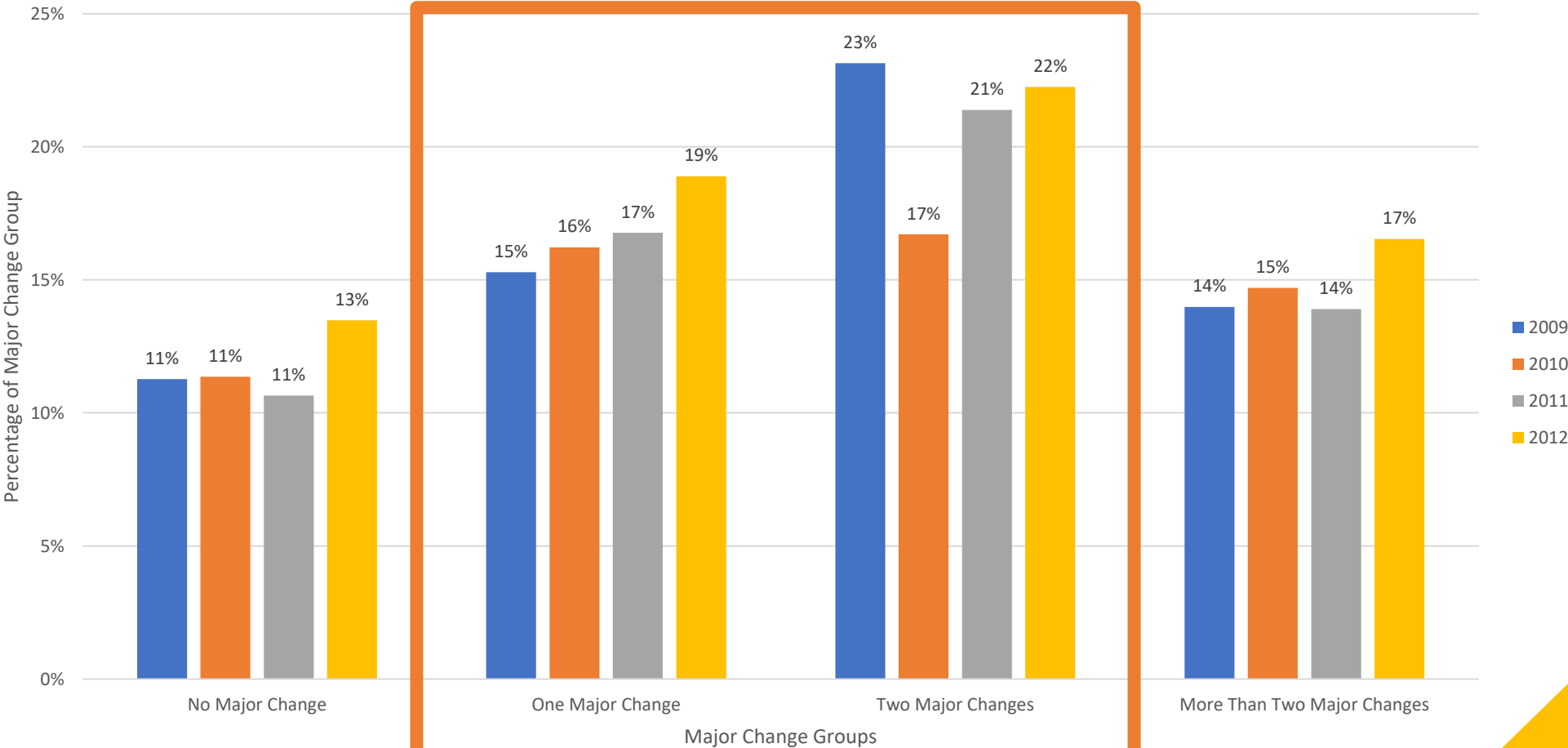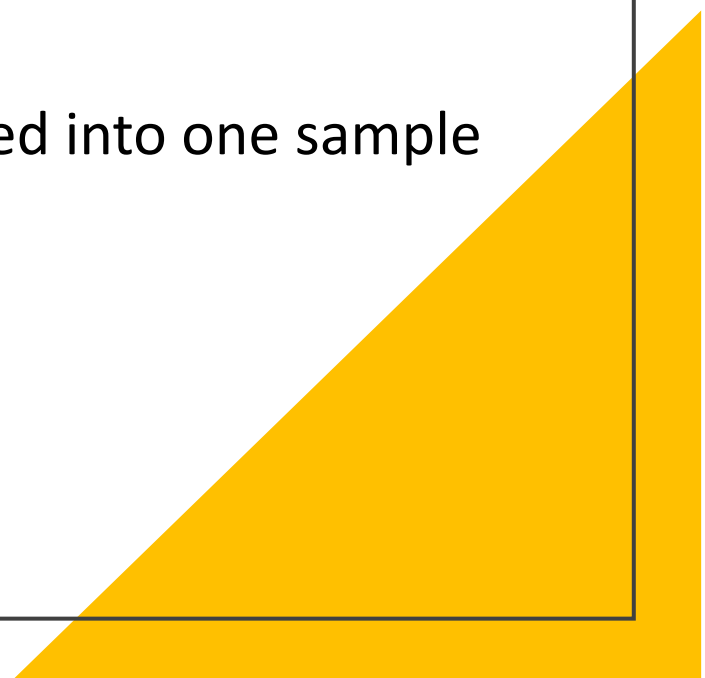
Six Year Graduation Rate Within Major Change Groups
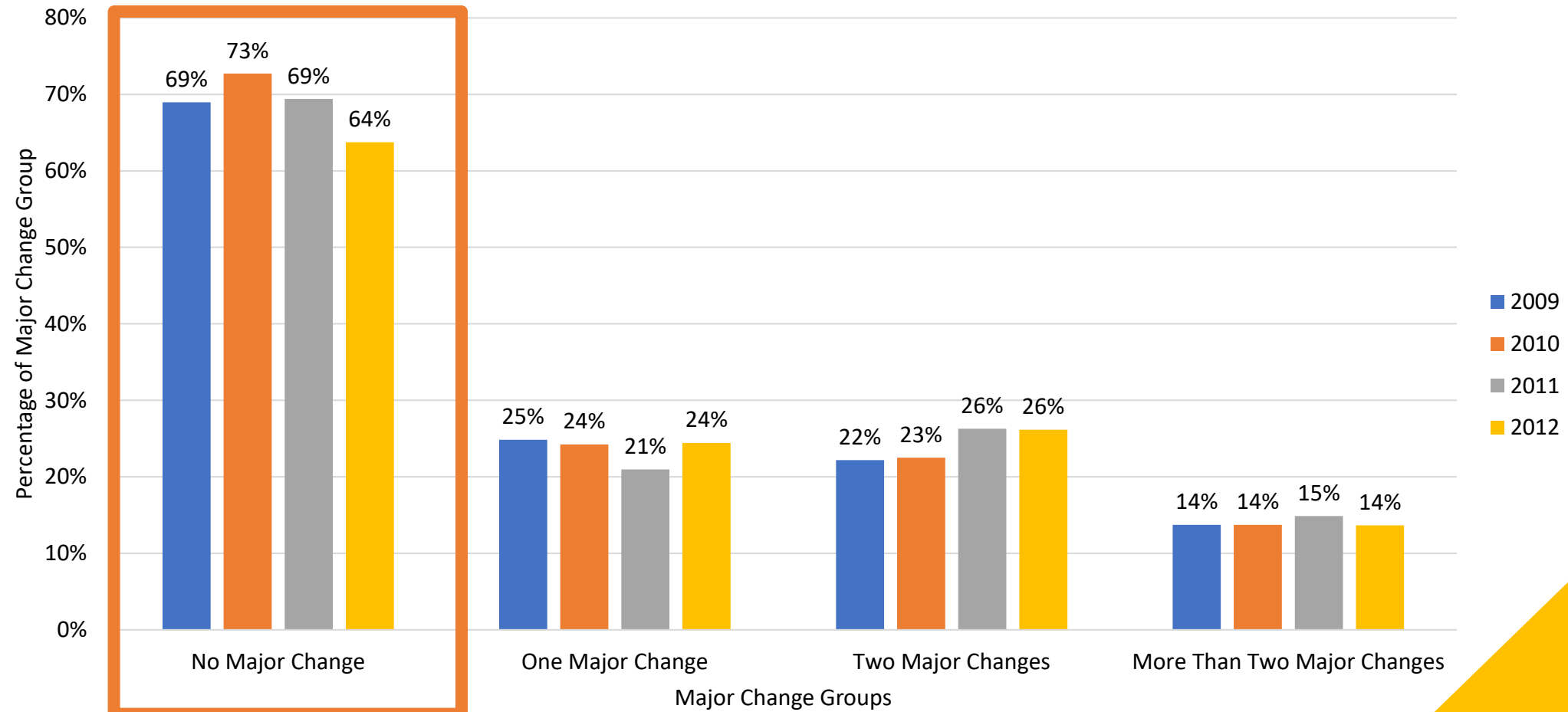
# Descriptive Statistics Conclusions

- There was no statistically significant difference in the student's academic preparation, and CSULB academic information between the four cohorts. Additionally, the major change rates were relatively stable through the cohort years.

- Based on these results, the four cohorts were combined into one sample before performing any analysis.

# Research Question

- Does major change have a significant effect on a student's graduation status, and can this result be combined with the student's academic and demographic information to identify significant factors that influence timely graduation?

# Does major change activity effect timely graduation?

**Methodology:**

- $\chi^2$ hypothesis test for Homogeneity of Proportions, $\alpha = 0.01$
- Phi Coefficient of correlation, at a significance level of $\alpha = 0.01$, between the binary variables major change and graduation

## **Does major change activity effect timely graduation?**

**Hypothesis Test Conclusions:**
- At a significance level of $\alpha = 0.01$ I conclude that there was a statistically significant difference between the proportions of students who graduated between at least one of the four major change groupings
- Post hoc testing revealed at a significance level of $\alpha_B = 0.0017$
  - Students who did not change majors graduated at a significantly different rate then students who changed majors
  - Students who changed majors once graduated at a significantly different rate than students who changed more than once

# How Are Major Change And Timely Graduation Related?

| Correlation between major change and graduated status | p-value |
|---|---|
| $r_\phi = 0.47$ | $p < 0.001$ |

| Major Change Group | Did not Graduate | Overall Graduation Rate |
|---|---|---|
| Did Not Change Majors | 69% | 31% |
| Changed Majors Once | 23% | 77% |
| Changed Majors Twice | 24% | 76% |
| Changed Majors More than Twice | 14% | 86% |

# Research Question

- Does major change have a significant effect on a student's graduation status, and can this result be combined with the student's academic and demographic information to identify significant factors that influence timely graduation?

# Dimensionality Reduction

- I began my study with 181 variables. After dropping variables with missing data I had 24 complete variables in the dataset.

- This is still relatively high dimensional data, and I want to reduce the dataset further without losing too much information

**Methodology:** Factor Analysis on the correlation matrix.

# Dimensionality Reduction :
# Factor Analysis on the Correlation Matrix

## Factor Loadings

| Variable Name | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|---|---|---|---|---|---|
| High School GPA | 0.98 | -0.19 | - | - | - |
| Overall CSULB GPA | 0.26 | -0.70 | 0.58 | 0.14 | 0.24 |
| Math Standardized Test Score | 0.20 | - | - | 0.92 | 0.32 |
| Reading Standardized Test Score | 0.20 | - | - | 0.31 | 0.85 |
| Eligibility Index Score | 0.86 | -0.16 | - | 0.31 | 0.35 |
| GE or Non-Major Credits | - | - | 0.84 | - | - |
| Major Credits | 0.12 | -0.25 | 0.58 | - | - |
| Total Number of DFW Courses | -0.19 | 0.82 | 0.21 | - | -0.13 |
| Total Number of WU Courses | - | 0.53 | - | - | - |
| Total Number of WE Courses | - | 0.16 | - | - | - |

# Dimensionality Reduction :
# Factor Analysis on the Correlation Matrix

**Factors and Their Associated Latent Variables**

| Factor Number | Latent Variable Explained |
|---|---|
| Factor 1 | Academic Preparation |
| Factor 2 | CSULB Academic Information |
| Factor 3 | Credits Taken |
| Factor 4 | Math Standardized Test Score |
| Factor 5 | Reading Standardized Test Score |

# Dimensionality Reduction: The Final Dataset

| Categorical Variables | Continuous and Discrete Variables |
|---|---|
| STEM Admission | Factor 1 - Academic Preparation |
| Student Gender | Factor 2 - CSULB Academic Information |
| Pell Eligibility | Factor 3 - Credits Taken |
| Local Admission | Factor 4 - Math Standardized Test Score |
| Minority Admission | Factor 5 - Reading Standardized Test Score |
| First Generation Student | Number of Major Changes |

## Modeling Time To Graduation

**Goals:** Determine significant variables that influence time to graduation

**Methodology:** Multinomial Logistic model. Variable selection will be performed using backward elimination at a significance of $\alpha = 0.05$

A multinomial logistic model is used to describe the relationship between predictor variables and a multi-level dependent variable whose outcomes are mutually exclusive. Outcomes are interpreted in relationship to the reference category.

In this study the outcomes are "Did not Graduate", "Four-Year Graduate", "Six-Year Graduate" and the reference category was set to "Four-Year Graduate"

# Modeling Time To Graduation: A Multinomial Logistic Model

After 1 fitting, the estimated coefficient for the predictor variable First Generation Status was found to be not significantly different from zero (p-value larger than 0.05), and all other predictor variables had a p-values less than 0.05.

2 final models were created:
1. The first model explained how the predictor variables increased a student's risk to not graduate compared to graduating in 4 years.
2. The second model explained how the predictor variables increased a student's risk to graduate in 6 years compared to graduating in 4 years.

# Modeling Time To Graduation: A Multinomial Logistic Model

Model 1: Comparing the risk of not graduating versus graduating in four years

$$\ln\left(\frac{P(Y = \text{ Did not Graduate})}{P(Y = \text{Four Year Graduate})}\right)$$

$$= 2.43 + 0.99(\text{STEM ADMISSION}) + 0.16(\text{STUDENT GENDER})$$

$$+ 0.12(\text{PELL ELIGIBILITY}) - 0.18(\text{LOCAL ADMISSION STATUS})$$

$$- 0.47(\text{MINORITY STATUS}) - 0.76(\text{FACTOR 1}) + 3.03(\text{FACTOR 2})$$

$$- 3.76(\text{FACTOR 3}) - 0.47(\text{FACTOR 4}) - 0.90(\text{FACTOR 5})$$

$$- 0.42(\text{NUMBER OF MAJOR CHANGES})$$

# Modeling Time To Graduation: A Multinomial Logistic Model

**Relative Risk Ratios for Did Not Graduate Versus Four Year Graduate**

| Variable Name | Y = Did Not Graduate vs Four Year Graduate |
|---|---|
| STEM Admission | 1.10 |
| Student Gender | 1.18 |
| Pell Eligibility | 1.14 |
| Local Admission | 0.84 |
| Minority Admission | 0.62 |
| Factor 1 - Academic Preparation | 0.47 |
| Factor 2 - CSULB Academic Information | 20.64 |
| Factor 3 - Credits Taken | 0.02 |
| Factor 4 – Math Standardized Test Score | 0.63 |
| Factor 5 - Reading Standardized Test Score | 0.41 |
| Number of Major Changes | 0.66 |

The CSULB Academic Variable is a linear combination of the number of courses the student did not pass and their GPA.

Positive values in this variable indicate the student performed poorly while enrolled

# Modeling Time To Graduation: A Multinomial Logistic Model

Model 2: Comparing the risk of graduating in six years versus graduating in four years

$$\ln\left(\frac{P(Y = \text{ Six Year Graduate})}{P(Y = \text{Four Year Graduate})}\right)$$

$= 1.01 + 0.36(\text{STEM ADMISSION}) - 0.36(\text{STUDENT GENDER})$

$+ 0.30(\text{PELL ELIGIBILITY}) + 0.21(\text{LOCAL ADMISSION STATUS})$

$- 0.01(\text{MINORITY STATUS}) - 0.22(\text{FACTOR 1}) + 0.79(\text{FACTOR 2})$

$+ 0.73(\text{FACTOR 3}) - 0.19(\text{FACTOR 4}) - 0.57(\text{FACTOR 5})$

$+ 0.12(\text{NUMBER OF MAJOR CHANGE}$

# Modeling Time To Graduation: A Multinomial Logistic Model

**Relative Risk Ratios for Six Year Graduate Versus Four Year Graduate**

| Variable Name | Y = Six Year Graduate vs Four Year Graduate |
|---|---|
| STEM Admission | 1.44 |
| Student Gender | 0.70 |
| Pell Eligibility | 1.36 |
| Local Admission | 1.24 |
| Minority Admission | 0.99 |
| Factor 1 - Academic Preparation | 0.80 |
| Factor 2 - CSULB Academic Information | 2.20 |
| Factor 3 - Credits Taken | 2.08 |
| Factor 4 – Math Standardized Test Score | 0.83 |
| Factor 5 - Reading Standardized Test Score | 0.56 |
| Number of Major Changes | 1.12 |

## Classifying Observations

**Goal:** Derive a classification algorithm that can be used to classify new observations into "Did Not Graduate", "Four-Year Graduate", or "Six-Year Graduate".

**Methodology:** Fishers Linear Multi-Population Discriminant Analysis

$\pi_1$ = "Did Not Graduate", $p_{\pi_1} = 0.35357$, $n_{\pi_1} = 2{,}406$

$\pi_2$ = "Four-Year Graduate", $p_{\pi_2} = 0.1611$, $n_{\pi_2} = 1{,}096$

$\pi_3$ = "Six-Year Graduate", $p_{\pi_3} = 0.4854$, $n_{\pi_3} = 3{,}303$

Note: I assumed equal misclassification costs.

# Classifying Observations: Fishers Linear Discriminant Analysis (LDA)

**Coefficients of Linear Discriminants**

| Variable Name | Coefficient LD1 | Coefficient LD2 |
|---|---|---|
| STEM Admission | 0.0533 | 0.4355 |
| Student Gender | −0.1157 | −0.2910 |
| Undeclared Admission | 0.1355 | 0.5943 |
| Pell Eligibility | 0.0662 | 0.2684 |
| Local Admission | 0.1112 | 0.2374 |
| Minority Admission | 0.1722 | −0.0856 |
| First Generation Student | 0.0130 | 0.0835 |
| Number of Major Changes | 0.2955 | 0.2159 |
| Factor 1 - Academic Preparation | 0.2198 | −0.2859 |
| Factor 2 - CSULB Academic Information | −0.7131 | 0.5219 |
| Factor 3 - Credits Taken | 1.0944 | 0.2711 |
| Factor 4 - Math Standardized Test Score | 0.0874 | −0.2126 |
| Factor 5 - Reading Standardized Test Score | 0.1205 | −0.5139 |
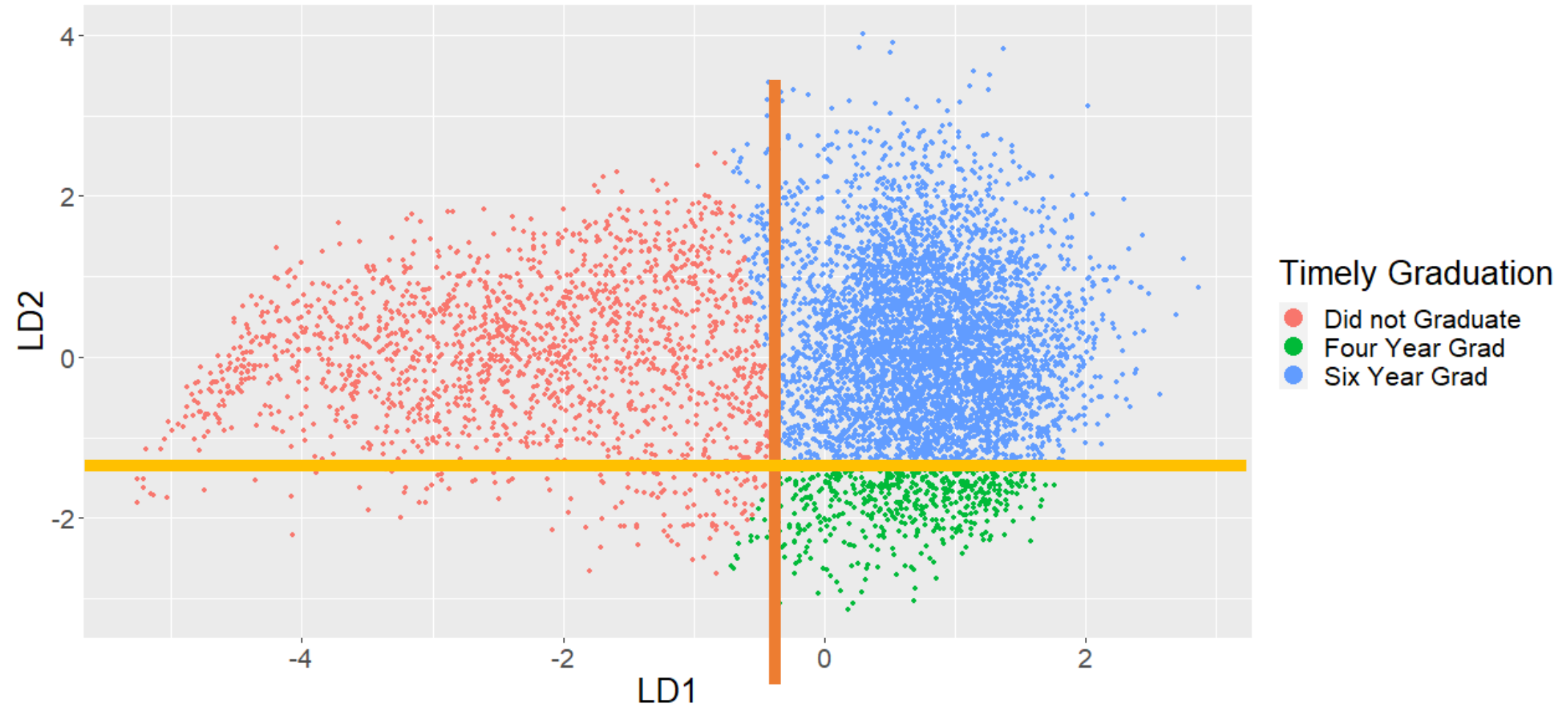
# Classifying New Observations: Fishers Linear Discriminant Analysis (LDA)

| Categorical Variables | Continuous and Discrete Variables |
|---|---|
| STEM Admission | 0 |
| Student Gender | 0 |
| Pell Eligibility | 1 |
| Local Admission | 0 |
| Minority Admission | 0 |
| First Generation Student | 0 |
| Factor 1 - Academic Preparation | 0.3834 |
| Factor 2 - CSULB Academic Information | -0.1500 |
| Factor 3 - Credits Taken | 0.0178 |
| Factor 4 - Math Standardized Test Score | 1.1365 |
| Factor 5 - Reading Standardized Test Score | 1.9898 |
| Number of Major Changes | 2 |
| True Classification | Six-Year Graduate |

| LD1 | LD2 |
|---|---|
| 1.2071 | 1.5173 |

# Classifying New Observations: Fishers Linear Discriminant Analysis (LDA)
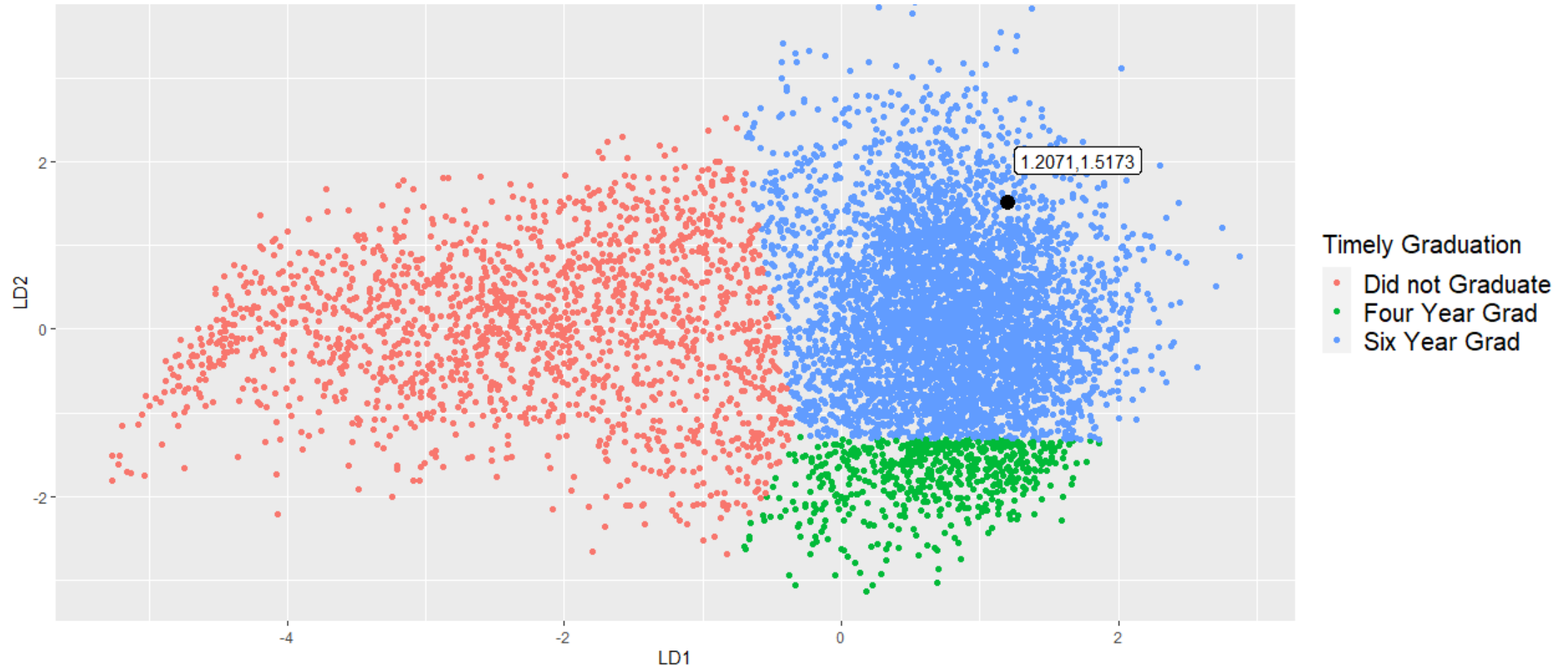
## Classifying Observations: Conclusion

- When classifying new observations, both the Multinomial Logistic model and Fishers Multipopulation LDA did a good job of separating students who did not graduate from those who did.

- Fishers Multipopulation LDA was better at classifying new observations than the Multinomial Logistic Model, however the Multinomial Logistic model was easier to interpret

- Based on these results, I concluded that a Multinomial Logistic model should be used to explain the relationships between the timely graduation and significant predictors, while Fishers Multipopulation LDA should be used to classify new observations.

# Conclusions

- **Research Question:** Does major change have a significant effect on a student's time to graduation and can those results be combined with the students' academic and demographic information to identify significant factors that influence timely graduation?

- Students who changed majors at least once graduated at a significantly different rate than students who did not change majors at a significance level of $\alpha = 0.01$

- The Phi Coefficient of Correlation showed a medium positive correlation between major change and graduation status. Students who changed majors were more associated with graduating than students who did not change majors.

# Conclusions

- At a significance level of $\alpha = 0.05$, the most significant factor that increased the student's risk of not graduating was the student's value in the latent variable CSULB Academic Information. This variable is a weighted linear combination of the students CSULB GPA, and the number of DFW, WU, and WE courses the student had while enrolled at CSULB.

- STEM Majors, Pell Eligible students, and local admission students are more likely to graduate as six-year graduates instead of four-year graduates. Students who graduate but perform poorly academically tend to graduate as six-year graduates.

# Conclusions

- Fishers LDA with 3 populations ("Did Not Graduate", "Four-Year Graduate", and "Six-Year Graduate") resulted in a classification model that correctly classified 75.42% of the training dataset and 74.65% of the testing dataset.

- The first linear discriminate (LD1) was used to separate students who did not graduate from those who did. The second linear discriminant (LD2) was used to separate four-and six-year graduates.

# Conclusions

- These results suggest that applying Fishers LDA to the student population and calculating each student's discriminant score each semester will provide meaningful information to Academic Advising.
  - Students will enroll at CSULB with a value of zero in both linear discriminants, and their discriminate scores can be updated at the end of each semester. If the student's LD1 score develops a negative trend, Academic Advising can contact these students and provide needed support to help retain these students.  The student's average LD2 score can be calculated to determine if a student is on track to graduate within four-years.
  - Additionally, the average LD1 and LD2 scores can also be calculated for each college and major within CSULB to create a profile of students within each major. Stakeholders can then focus on colleges and majors where students traditionally have negative LD1 scores and positive LD2 scores.

# Limitations of the Study

- Due to the available data, the Did Not Graduate variable is not well defined.
    - Students have multiple ways to exit before finishing their degree:  students can transfer to a new college to complete their degree, or they could drop out and not complete their degree.
- Due to the data available on the students who exited and the exclusion of students who enrolled in a Winter or Summer courses the four-year graduation rate reported in this study might be lower than the actual graduation rates.
- There is a large sample size imbalance between the graduation statuses:
    - 16% of the sample were four-year graduates
    - 48% of the sample were six-year graduates
    - 35% of the sample did not graduate
- Variables in this study are a combination of pre-admission and post-admission data.
    - It might be beneficial to create semester-based variables which can be used for early detection.

# Suggestions for Future Research

- Future research should study if there is a statistically significant relationship between Summer or Winter semester enrollment and timely graduation.

- On October 26th , 2020 the LA Times published an article which discussed Graduation Initiative 2025 and noted that low-income and underrepresented minority students are graduating at "rates of about 10 percentage points lower" than students not in those groups.  Future research should restrict the sample to these students and determine if there are specific risk factors for minority or low-income students.

- From Fall 2020 to Fall 2024, the University of California system will not use either the SAT or ACT test scores in their admission process (Gordon, 2020).  The models should be re-worked without these predictors to prepare for the CSU system to follow

## Suggestions for Future Research

- Future research should examine starting and ending majors for students who change majors and do not graduate. It might be beneficial for Academic Advising to have this information to share with students who want to change majors.

- Should CSULB modify its admission criteria to ensure that more students graduate college or graduate college in 4 years? Or, if we keep the current admission criteria the same, what can the institution do to support those who are at risk of not graduating?

# Questions

- Does anyone have any questions that I can answer about my research?

- Please use the raise hand feature in the participants tab so we can avoid talking over each other or type your question in chat.