

# **Development of Logistic Regression Models to Predict Graduation in Higher Education and STEM Majors**

By

Ujwala Paladi, B.Sc.

A Thesis Submitted in Partial Fulfillment  
Of the Requirements for the Degree  
Master of Science  
Major Subject: Mathematics

West Texas A&M University

Canyon, Texas

August, 2017

Approved:

---

Thesis Committee Chairman

---

Date

---

Thesis Committee Member

---

Date

---

Thesis Committee Member

---

Date

---

Department Head

---

Date

---

Dean, Academic College

---

Date

---

Graduate School Dean

---

Date

## **ABSTRACT**

The topic of this thesis is the analysis of factors influencing the retention of undergraduate level students at West Texas A&M University (WTAMU) with a special emphasis on STEM majors. The subjects of this study are first time college students, enrolled at WTAMU in Fall 2010, Fall 2011 and Fall 2012 that have a declared first major in a STEM field. Through Logistic regression analysis, factors influencing student retention and retention to graduation are identified. Identified factors that influence undergraduate student retention and graduation at WTAMU are math and science self-confidence, study habits, HSGPA, class percent, mothers' education, and senior year grades. Factors identified that influence retention to graduation of first year STEM majors at WTAMU are percentile of transfers, desire to finish, self-reported college prep activity, highest degree sought, sociability, distance from campus, HSGPA, class percent, Major code, fathers education, and work.

## **AKNOWLEDGEMENTS**

I thank my mentor Dr. Lockwood for her invaluable scientific advice, technical help, and encouragement throughout my thesis. Dr. Lockwood has been my role model and backbone throughout my journey at WTAMU. Words can't express my gratitude towards her for support through rough patches and TA position.

I express sincere thanks to Dr. Seth for his valuable suggestions and his constructive criticism during discussions and writing process. He is one of the nicest persons I have ever met, and thank him for financially helping me out with a grader position.

Maria and Amanda, I always cherish the time I spent with you, and huge thank you for helping me out on several occasions.

I thank my husband Ramakrishna for his encouragement without his help and support this thesis would not come to fruition.

I extend my sincere, heartfelt thanks to Padma aunty for her emotional support and love. She is one of the strongest and brave ladies I have ever met. I learned how to be open minded and how to express myself without fear from her.

I extend my sincere, heartfelt thanks to my sister Sreelatha; she is my role model of helping nature. I thank her for good food, cordial relationship, taking care of Vishnu, and for encouragement.

I thank my mother for her emotional support and being there for me to lean on. Her presence made my life easier during tough times.

I thank my stress buster son Vishnu, Srilekha and Manus for behaving nice and understanding me.

Also take this opportunity to thank my extended family and friends

Last but not least, one person that would have been happiest at this moment is my dad. He is no more to his daughter in a graduation gown, but I hope he is watching me from the heavens, and I dedicate this thesis to my loving dad Narasimha Rao Paladi.

Ujwala Paladi

August 2017

## **TABLE OF CONTENTS**

<b>ABSTRACT .....</b>	<b>iii</b>
<b>AKNOWLEDGEMENTS .....</b>	<b>iv</b>
<b>TABLE OF CONTENTS .....</b>	<b>vi</b>
<b>LIST OF TABLES .....</b>	<b>vii</b>
<b>LIST OF FIGURES .....</b>	<b>viii</b>
<b>CHAPTER I: INTRODUCTION.....</b>	<b>1</b>
<b>CHAPTER II: METHODOLOGY .....</b>	<b>4</b>
<b>CHAPTER III: FACTORS AFFECTING RATE OF GRADUATION .....</b>	<b>13</b>
<b>CHAPTER IV: FACTORS AFFECTING RATE GRADUATION IN STEM MAJORS.....</b>	<b>57</b>
<b>REFERENCES .....</b>	<b>96</b>

## LIST OF TABLES

Table 1: Single logistic regression analysis output between undergraduate student graduation and independent variables.....	44
Table 2: List of factors and their p-values that have been dropped from regression analysis.....	52
Table 3: Final list of factors influencing rate of graduation .....	55
Table 4: Single logistic regression analysis output between STEM student graduation and independent variables.....	80
Table 5: Final list of factors with significant effect on student retention and graduation in higher educations with STEM majors.....	88

## LIST OF FIGURES

Figure 1: DISTRIBUTION OF STUDENT POPULATION BASED ON THEIR SUCCESS OR FAILURE .....	6
Figure 2: STUDENT POPULATION DISTRIBUTION BASED ON END RESULT ...	14
Figure 3: PIE CHART OF GENDER .....	15
Figure 4 : PIE CHART OF RACE.....	16
Figure 5: PIE CHART OF MAJOR CODE.....	17
Figure 6: PIE CHART OF DORM .....	18
Figure 7: PIE CHART OF PELL.....	19
Figure 8: PIE CHART OF FATHERS EDUCATION .....	20
Figure 9: PIE CHART OF MOTHERS EDUCATION .....	21
Figure 10: PIE CHART OF SENIOR YEAR GRADES .....	22
Figure 11: PIE CHART OF SELF REPORTED TIMING OF DECISION .....	23
Figure 12: PIE CHART OF WORK .....	24
FIGURE 13: BOX PLOTS OF MATH AND SCIENCE CONFIDENCE PERCENTILE .....	25
Figure 14: BOX PLOTS OF DESIRE TO FINISH PERCENTILE .....	26
Figure 15: BOX PLOTS OF PARENT EDUCATION PERCENTILE .....	27
Figure 16: BOX PLOT OF RECEPTIVITY TO ACADEMIC ASSISTANCE.....	28
Figure 17: BOX PLOTS OF RECEPTIVITY TO FINANCIAL GUIDANCE.....	29
Figure 18: BOX PLOTS OF RECEPTIVITY TO PERSONAL COUNSELING .....	30
Figure 19: BOX PLOT OF RECEPTIVITY TO SOCIAL ENRICHMENT.....	31
Figure 20: BOX PLOTS OF ACADEMIC STRESS PERCENTILE.....	32
Figure 21: BOX PLOTS ATTITUDE TOWARD EDUACTORS PERCENTILE .....	33
Figure 22: BOX PLOTS OF FAMILY EMOTIONAL SUPPORT PERCENTILE .....	34
Figure 23: BOX PLOTS OF SENSE OF FINANCIAL SECURITY PERCENTILE.....	35
Figure 24: BOX PLOTS OF SELF REPORTED COLLEGE PREP PERCENTILE .....	36



Figure 25: BOX PLOTS OF COLLEGE PREP PERCENTILE .....	37
Figure 26: BOX PLOTS OF HSGPA .....	38
Figure 27: BOX PLOTS OF OPINION TOLERANCE PERCENTILE .....	39
Figure 28: BOX PLOTS OF STUDY HABITS PERCENTILE .....	40
Figure 29: BOX PLOTS OF HIGH SCHOOL GPA .....	41
Figure 30: BOX PLOTS OF MAX ACT SAT SCORE .....	42
Figure 31: BOX PLOTS OF CLASS PERCENT .....	43
Figure 32: DISTRIBUTION OF GRADUATES BASED ON MAJOR .....	57
Figure 33: PIE CHART OF GENDER .....	59
Figure 34: PIE CHART OF RACE.....	60
Figure 35: PIE CHART OF PELL.....	61
Figure 36: PIE CHART OF FATHERS EDUCATION .....	62
Figure 37: PIE CHART OF MOTHERS EDUCATION .....	63
Figure 38:PIE CHART OF MAJOR.....	64
Figure 39: PIE CHART OF TRANSFER DESIRE .....	65
Figure 40: PIE CHART OF SENIOR YEAR GRADES .....	66
Figure 41: PIE CHART OF WORK .....	67
Figure 42: BOX PLOT OF MAX ACT & SAT SCORE .....	68
Figure 43: BOX PLOT OF MATH AND SCIENCE CONFIDENCE PERCENTILE ....	69
Figure 44: BOX PLOT OF HIGH SCHOOL GPA .....	70
Figure 45: BOX PLOTS OF COLLEGE PREP PERCENTILE .....	71
Figure 46: BOX PLOTS OF CLASS PERCENT .....	72
Figure 47: BOX PLOTS OF SOCIABILITY PERCENTILE .....	73
Figure 48: BOX PLOTS OF DISTANCE FROM CAMPUS.....	74
Figure 49: BOX PLOTS OF CAREER CLOSURE PERCENTILE .....	75
Figure 50: BOX PLOT OF TRANSFER PERCENTILE .....	76
Figure 51: BOX PLOTS OF SELF REPORTED COLLEGE PREP .....	77
Figure 52: BOX PLOT OF DESIRE TO FINISH PERCENTILE .....	78

Figure 53: BOX PLOT OF RECEPTIVITY TO ACADEMIC ASSISTANCE.....	79
--	----

## CHAPTER I: INTRODUCTION

One in three students enrolled in a college or university may not return for the sophomore year. A number of universities have identified several factors influencing student drop outs including personal and academic difficulties, part time status, choosing the wrong subject/program/university, no clarity in career goals, workload demand and poor quality faculty [1]. A detailed analysis of all factors is imperative to encourage student retention in higher education institutions.

STEM (Science, technology, engineering and mathematics) is a term that refers to the academic disciplines associated with the sciences, technology, engineering and mathematics. Students from STEM majors are in high demand in many industries and careers in STEM areas are thought of as a backbone of the American economy [2]. However, the percentage of American students pursuing higher education in STEM fields is low and does not meet the current demand. President Obama requested educators to develop programs to increase the number of students with STEM skill sets [3]. There is a general prevailing consensus that preventing students from switching from STEM majors to non-STEM majors is easier [4]. However, in spite of extensive high school preparation in STEM courses, more than half of the college or university STEM students choose to switch to liberal arts majors. Understanding factors influencing retention in STEM majors is challenging, and it is the need of the hour.

The subjects of this investigation are student who joined WTAMU's undergraduate programs. Logistic regression is implemented to investigate factors that impact student retention to graduation and student retention specifically to graduation in a

STEM field. The first query is to identify the internal and external factors associated with a first-time in college STEM major at a regional institution being retained to obtaining a higher education degree, at the two-year or four-year level. The second query is given a first-time in college STEM major is retained to degree completion at a higher education institution, what are the internal and external factors associated with retention to graduation specifically in a STEM field. Researchers at West Texas A&M University have been investigating factors that impact retention of first year students electing to major in a STEM discipline since 2014. The study population consists of three cohorts of first-time in college STEM majors from fall 2012, fall 2013, and fall 2014 who participated in the Noel-Levitz student attitude survey conducted with all first time students during these years. Prior researchers investigated the following research question using this study population.

What internal and external factors impact retention to graduation of first year STEM majors at a regional institution?

The methodology used to investigate this research question was logistic regression. A logistic regression model with an outcome variable indicating retained to graduation in a STEM field was successfully developed using a database of predictor variables that included internal and external factors. However, interpretation of regression coefficients indicated many seemingly contradictory statements, particularly found in interaction terms. For example, the higher the ACT score, the less likely a student was to be retained. As a result of these results, this researcher hypothesized that perhaps the study population consisted of two distinct populations of students. Population Subset One might consist of students with substantial external challenges, perhaps prohibiting

retention at the university level, much less a STEM major. Population Subset Two could be students with fewer external challenges whereby staying at the university was not the issue, but staying in a STEM field presented different challenges. This research hypothesizes that these two populations have very unique and different challenges that must be identified. This research will address the following research questions.

Research Question 1: What are the internal and external factors associated with a first-time in college STEM major at a regional institution being retained to graduation at an institution of higher education, either two-year or four-year?

Research Question 2: Given a first-time in college STEM major is retained to graduation at a higher education, what are the internal and external factors associated with retention to graduation in a STEM field?

### **Key Terms**

Dichotomous: A Categorical variable with two conceivable classes

Logistic Regression: A statistical method for examining a dataset in which there is at least one free factor that decide a result. The result is measured with a dichotomous variable (in which there are just two conceivable results).

P-value: A p value of 0.5 indicates that there is a 50-50 chance that the findings of the study are significant. A p value of 0.05 (the value usually used to provide evidence hypothesized results or models are statistically significant) implies that there is a 5% chance that the results of the study happened by chance.

## **CHAPTER II: METHODOLOGY**

The subjects of this investigation are students that one joined First time in college freshman electing to major in a STEM field in fall 2010, fall 2011 and fall 2012 at WTAMU's undergraduate programs. This research analyzes factors that impact student decisions during higher education, namely decisions to stay in school or not, and/or journey to graduation in a STEM or non-STEM field. First, we investigate academic, social, fiscal and psychological factors that impact these students' decision to stay in higher education or drop-out. Second, we focus on these students who chose to remain at a four year institution and demystify factors affecting their graduation in either a STEM or NONSTEM degree programs.

### **DATA AND COLLECTION PROCESS**

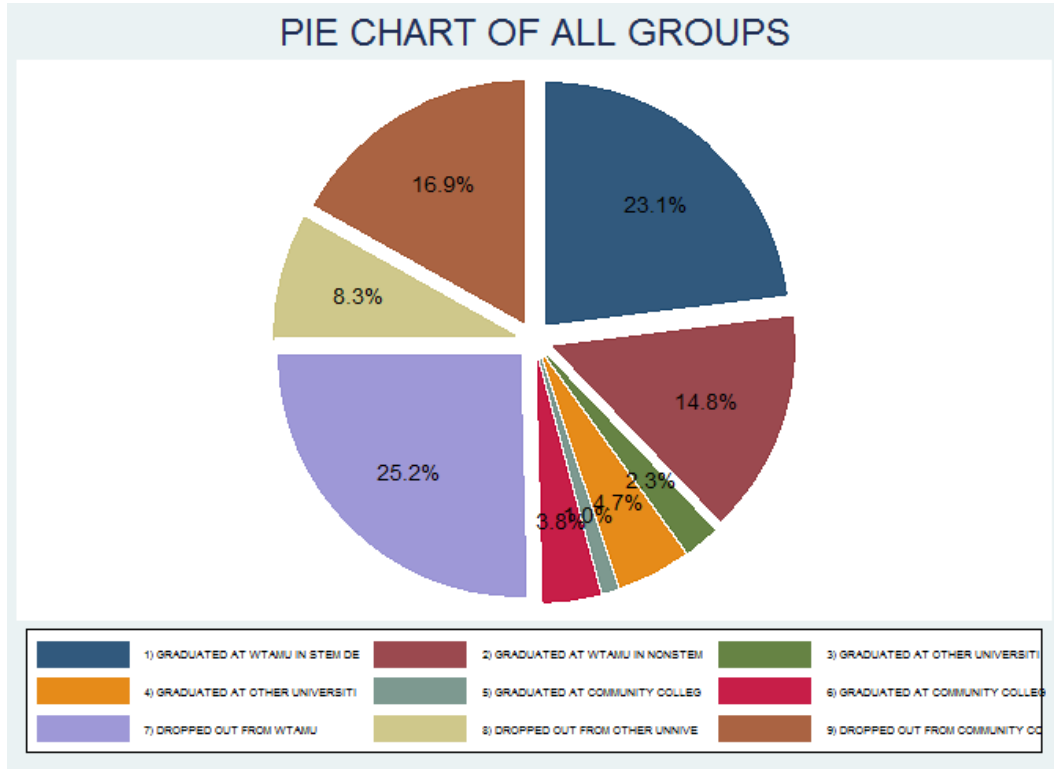
WTAMU partnered with an outside analysis firm, Noel-Levitz, to research and collect data, analytics and assessment of enrollment services, student success services, fund raising services and an array of organizational wide services. The majority of the data used for this study was collected by a Noel-Levitz survey of 100 questions (academic, social, financial, and psychological) involving fall 2010, fall 2011 and Fall 2012 recruits. A copy of the survey is incorporated into Appendix I. Detailed procedure on data collection and data processing on the Noel-Levitz survey can be found at the Noel-Levitz site: <https://www.noellevitz.com>. Additional background information on each student, such as ACT/SAT scores, class rank, demographics, financial need, etc. was provided by the West Texas A&M Office of Institutional Research and the Office of Admissions.

This researcher further partnered with the National Student Clearing House through the WT Office of Institutional Research to obtain information regarding students who left WT, specifically transfer students journeys to graduation or drop out at other institutions. National Student Clearing house is a nonprofit and nongovernmental organization and a leading provider of educational reporting, data exchange, verification, and research services to help educational institutions to meet compliance, administrative, student access, accountability, and analytical Needs. The student data obtained from the Noel-Levitz survey alongside other qualitative and quantitative information gathered by the Office of Institutional Research and the Office of Admissions at WTAMU, was employed to create a database of factors that could influence retention to graduation of first year STEM students at WTAMU. An Excel data file was created and maintained with all the data points, 1004 students were included in this study. There were 568 male students and 436 female students.

First, I divided the data into 9 groups. They are given below:

- 1) WTAMU graduates with STEM major (23.1%).
- 2) WTAMU graduates with NONSTEM major (14.8).
- 3) Other university graduates with STEM major (2.3%).
- 4) Other university graduates with NONSTEM major (4.7%).
- 5) Community college graduates with TECH program (0.9%)
- 6) Community college graduates with associative degree (3.8%)
- 7) Drop outs from WTAMU (25.2%)

- 8) Drop outs from other universities (8.3%)
- 9) Drop outs from community colleges (16.9%).



**FIGURE 1: DISTRIBUTION OF STUDENT POPULATION BASED ON THEIR SUCCESS OR FAILURE**

About half (50.4%) of the students dropped out without any degree, while 44.8% of graduated with an undergraduate degree, and 4.8% graduated with a college degree.



## **Dependent variables**

Based on the research questions two categories of outcome dependent variables were created. Outcome variable 1, RETAINED, identified whether or not the student decided to remain enrolled in higher education. As subjects were tracked utilizing the Clearinghouse, researchers could identify if a student remained at WT, left WT, and enrolled at another four-year institution, enrolled at another two-year institution, or never returned to higher education. If a student remained at WT or enrolled in an institution of higher education, four year or two year, prior to fall 2016, they were considered Retained to Higher education. If the student displayed no further enrollment at WT or any other institution of higher education then, they were considered Not Retained. The variable RETAINED is therefore a dichotomous outcome variable. Outcome variable 2, STEM, was considered for only those subjects who remained enrolled in higher education. If the student graduated with a STEM degree from a four year institution of higher education, or was still enrolled with a STEM major, then the student was considered retained in a STEM degree. If the student graduated from a four year or two year institution of higher education in a non-STEM field, then the student was considered Not Retained. The STEM outcome variable is then also a dichotomous outcome variable.

## **Categorical Predictor variables**

The following are the categorical predictor variables considered for this study. Each variable has two or more categorical outcomes which were coded with appropriate numerical values for further exploratory and statistical analysis.

**Gender**

Two categories Male (0) and Female (1).

**Race**

Race was divided into 4 categories. The white students (1), the Hispanic students (2), the African American students (3), and the students identified as multi-ethnic or other (4).

**Degree sought**

Degree sought was divided into three categories. Bachelors (1), professionals (2), and post graduate (3).

**Transfer desire**

Transfer desire indicates a student response to a direct question on the Noel-Levitz survey as to the student's desire to transfer. The question had two categories, Yes and No. Transfer desire No (0), and yes (1).

**Senior year grades**

On the survey, the students were asked to self-report their senior year grades. Student responses were divided into three categories. A-average (1), B+ average (2), B average and C+ average (3).

**Work**

Work requested students identify how many hours each week they intended to work, if at all, while enrolled in their first year in college. Outcomes were initially divided into six categories. Student did not anticipate working (0); a student anticipating

working 1 to 10 hours a week (1); a student anticipating working 11 to 20 hours a week (2); a student anticipating working 21 to 30 hours a week (3); and a student anticipating working 31 to 40 hours a week (4).

### **Dormitory**

This variable indicated whether or not a student lived on campus in a dormitory during their first semester. The response was coded 0 if the student was not living in university dormitory, and 1 if the student was living in university dormitory.

### **Mother's education and father's education**

This variable represented the student response on the Noel-Levitz survey as to the combined education of their parents. The coding was as follows: the mother's or father's greatest achieved level of training was some primary school (0), some secondary school however no certificate (1), a secondary school confirmation or identical (2), one to three years of post-secondary school (3), a four year certification (4), a graduate degree (5), and an expert degree (6).

### **Self-reported timing of decision**

This variable identified the timing of the student's decision to attend college and was divided into three categories: few days before (1), few weeks before (2), and few months before (3).

### **Pell**

A student's eligibility for a Pell Grant was selected as a measure of the student's level of financial need. If student was eligible for a Pell grant, this variable was coded as 1. If a student was not eligible for a Pell Grant then it was coded as 0.

## **Major code**

Major code refers to the student's first-time in college major as determined by the twelfth class day of their first fall semester. This variable was coded as six categories initially. Categories may later be combined in order to prevent statistical modelling issues: Mathematics (1), Computer Science (2), Engineering (3), Pre-professional & Biology (4), Chemistry & Physics (5) and Agricultural Sciences (6).

## **Continuous variables**

The following is a list of continuous variables used as factors in this research. The Noel-Levitz survey included numerous survey questions which combined to form quantitative scale variables. These scales were then transformed to percentiles. The percentiles included as factors in this research from the Noel-Levitz survey are as follows: transfer percentile (indicates a student's desire to transfer from WT); desire to finish percentile; receptivity to academic assistance; receptivity to career counseling; self-reported college prep percentile (indicates a student's self-assessment of their preparation for college level work); math and science confidence percentile; career closure percentile (indicates a student's security in their current career choice); college prep percentile (based on student responses to typical college preparation coursework and activities); receptivity to financial guidance; receptivity to personal counseling; receptivity to social enrichment; academic stress percentile; attitude toward educators percentile; family emotional support percentile; sense of financial security percentile; verbal confidence percentile; intellectual interest percentile; sociability percentile; desire to transfer percentile, opinion tolerance percentile; and study habits percentile. Other Noel-Levitz percentile variables are highest degree sought percentile

e and high school grades percentile. Further continuous variables were obtained from the WT Office of Admissions and Institutional Research. These include: Max ACT and SAT score (maximum of ACT or SAT if both taken); distance from campus (distance of permanent address from Canyon, Texas); high school GPA; class percentile; age.

### **Logistic Regression**

The goal of a statistical model is “to find the best fitting and most tight, clinically interpretable model to describe the relationship between an outcome (dependent or response) variable and a set of independent (predictor or explanatory) variables. Logistic regression is used to predict an outcome variable, which is categorical or binary, from a collection of predictor variables, which can be continuous or categorical. In consideration of the research questions, two models need to be developed. First a model to investigate what factors effect staying in school and given stayed in school what factors influence remaining in STEM.

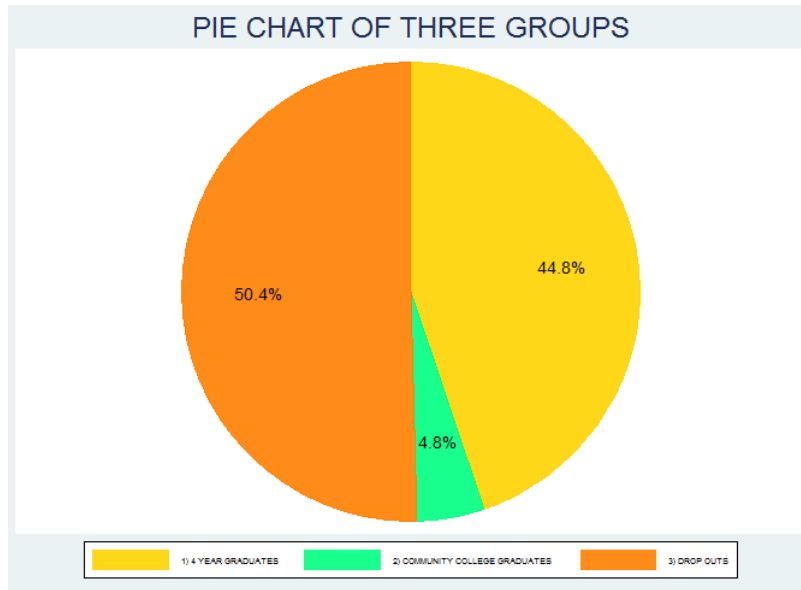
### **Analytic Design**

STATA ® (Release 14) was employed for data processing, graphing, analysis and interpretations. Logistic regression was implemented to analyze for the data. The principle behind the method implemented is backward elimination. An Initial regression model was constructed using a logit model to relate the dependent variable and independent variables. Then with all significant independent variables, logistic regression was performed to identify and remove the least significant independent variable. This process is repeated until the remaining variables are all significant independent variables. The final regression model was constructed using all significant predictors. A threshold

of 0.25 ( $\alpha = 0.25$ ) was used to identify all independent variables with significant association.

### **CHAPTER III: FACTORS AFFECTING RATE OF GRADUATION**

The first objective of this study was to identify factors that impact retention to graduation of a first time in college STEM major at WTAMU. This chapter will present the exploration and analysis related to answering the first research question. What factors impact retention to graduation in any degree program? However, remember, the student population of this study are in a STEM degree program. We will first present the exploration of the outcome variable, Graduation, indicating the student graduated from higher education in some manner, versus all predictor variables. Then the formal development of a logistic regression model for Graduation, based on these predictors will be developed.



**FIGURE 2: STUDENT POPULATION DISTRIBUTION BASED ON END RESULT**

**Inference:** About half (50.4%) of the students dropped out without any degree, while 44.8% of graduated with an undergraduate degree, and 4.8% graduated with a college degree.

### Exploration

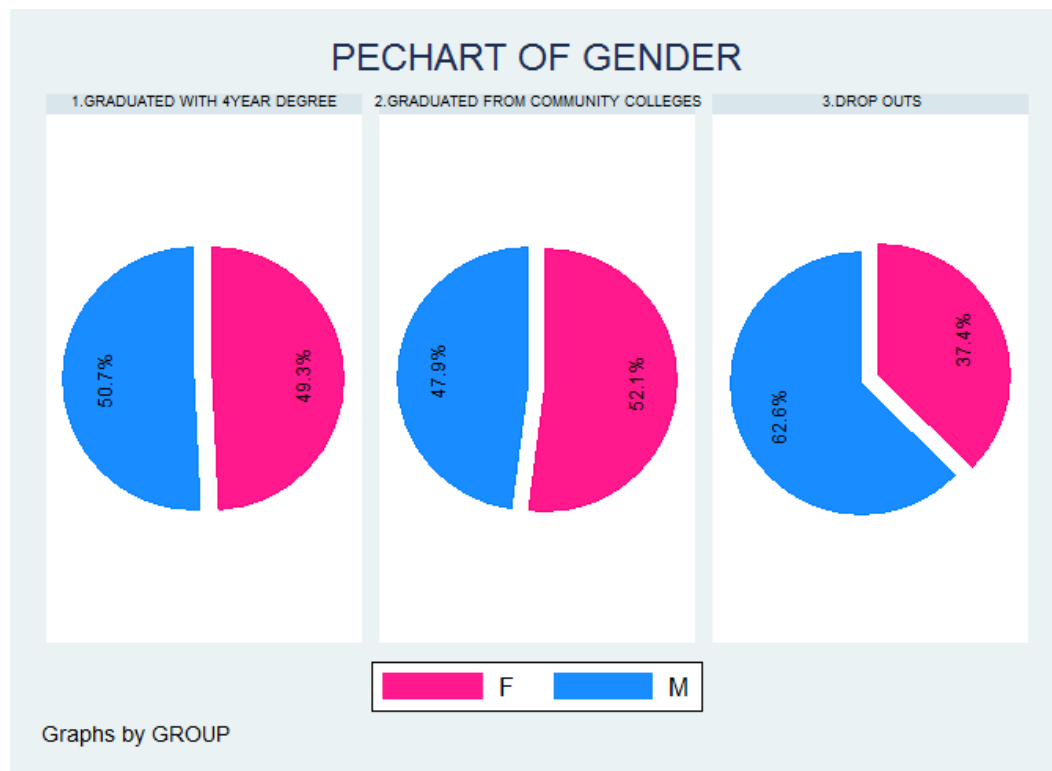
Some exploratory graphs were prepared before building the statistical model to guide the process and determine how to proceed. The exploration begins with identification of important factors affecting student success in 4-year degree programs or community college degree programs.

The following pie charts and Box and whisker plots address the 3 primary groups:

1) Students that graduated with a 4 year degree, 2) Students that ultimately graduated from a community college or 3) Students that ended dropping out of the higher education

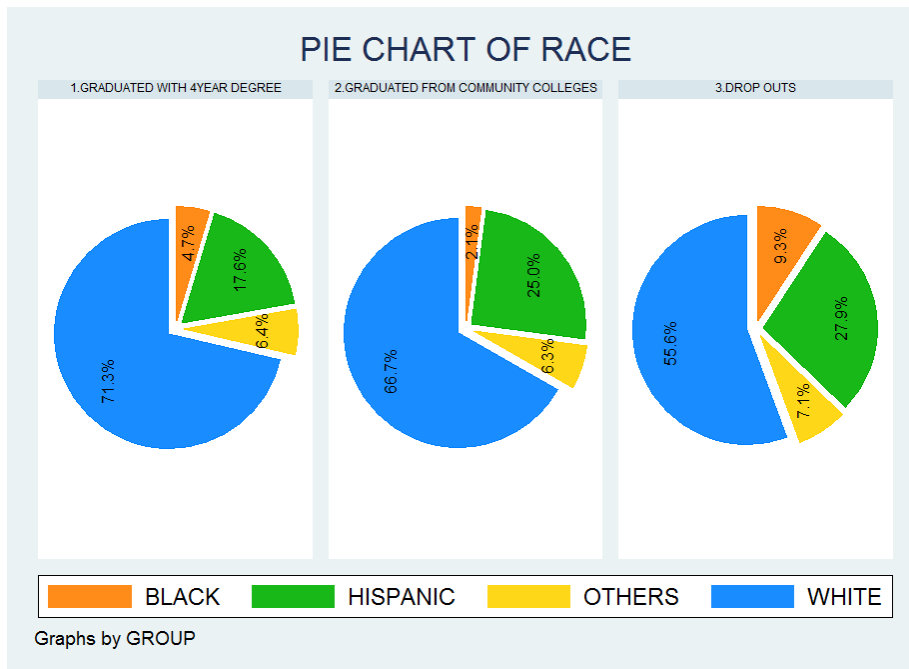


all together. For each of the variables the pie charts then indicate the proportions of students within the primary categories that satisfy the variables, options for example, the pie charts for the race variable, Figure 2 illustrates the proportion of students in 1), 2), and 3) of each race, example of the students that graduated from a community college 66.7% were whites, 25% Hispanic, 2.1% were blacks and 6.3% were other.



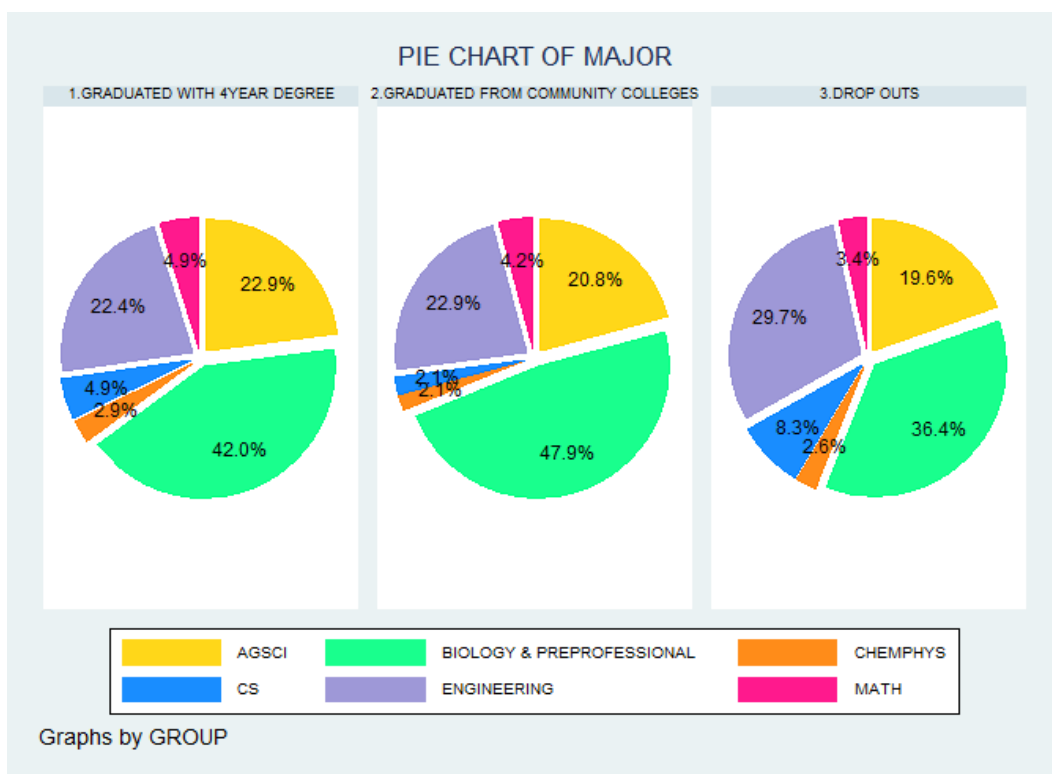
**FIGURE 3: PIE CHART OF GENDER**

**Inference:** Figure 3 displays the exploration of retention to graduation by gender. Of students that graduated with a 4 year degree or from community colleges, approximately equal numbers were male or female. However of those that dropped out, a higher proportion were males, 62.6%, than female, 37.4%.



**FIGURE 4 : PIE CHART OF RACE**

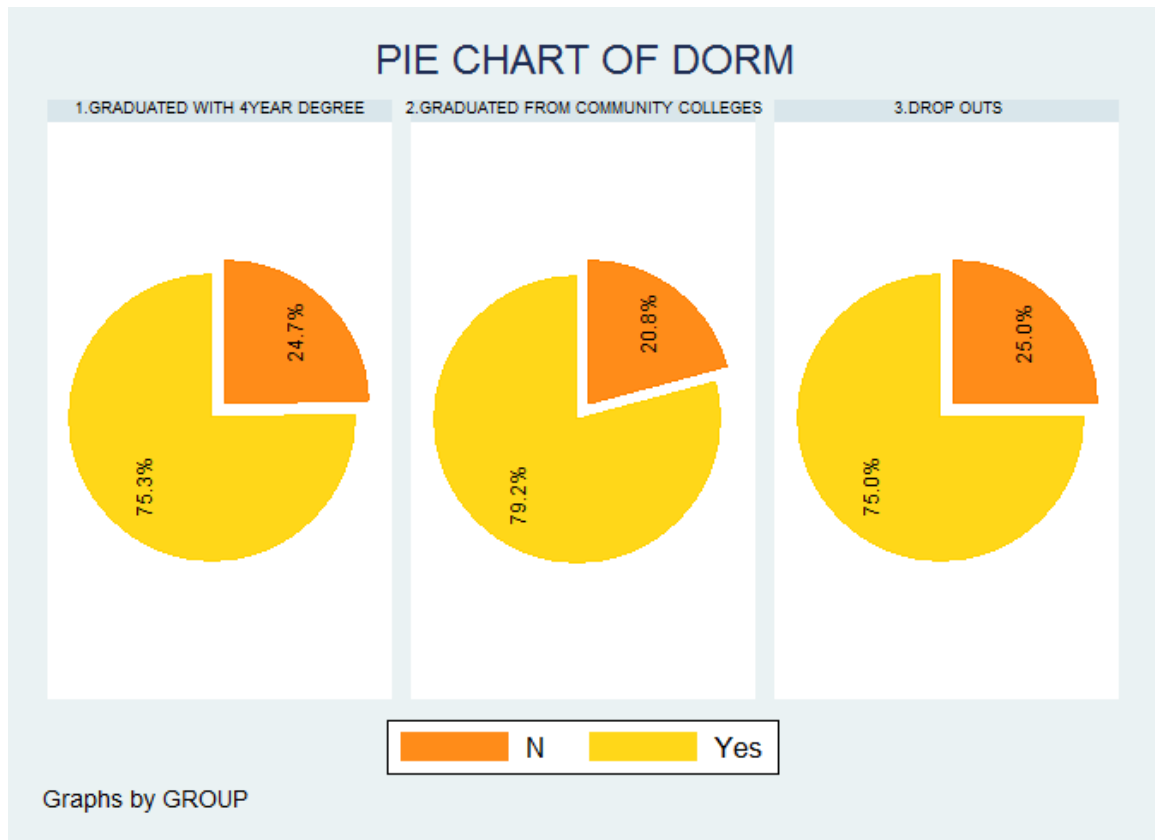
**Inference:** Figure 4 displays the racial distribution of each identified subgroup, four year graduates, two year graduates and drop-outs. Racial proportion of both 4-year degree graduates and community college graduates is similar, but proportion of whites in drop out lower.



**FIGURE 5: PIE CHART OF MAJOR CODE**

- 1) Major distribution among 4-year graduates: 4.9% (math), 4.9% (CS); 22.4% (engineering), 42% (Biology & pre professional), 2.9% (chemphys), and 22.9% (Agsci).
- 2) Major distribution among community college graduates: 4.2% (math), 2.1% (CS), 22.9% (engineering), 47.9% (Biology & pre professional), 2.1% (chemphys), and 20.8% (Agsci).
- 3) Major distribution among drop outs: 3.4% (math), 8.3% (CS), 29.7% (engineering), 36.4% (Biology & pre professional), 2.6% (chemphys) and 19.6% (Agsci)

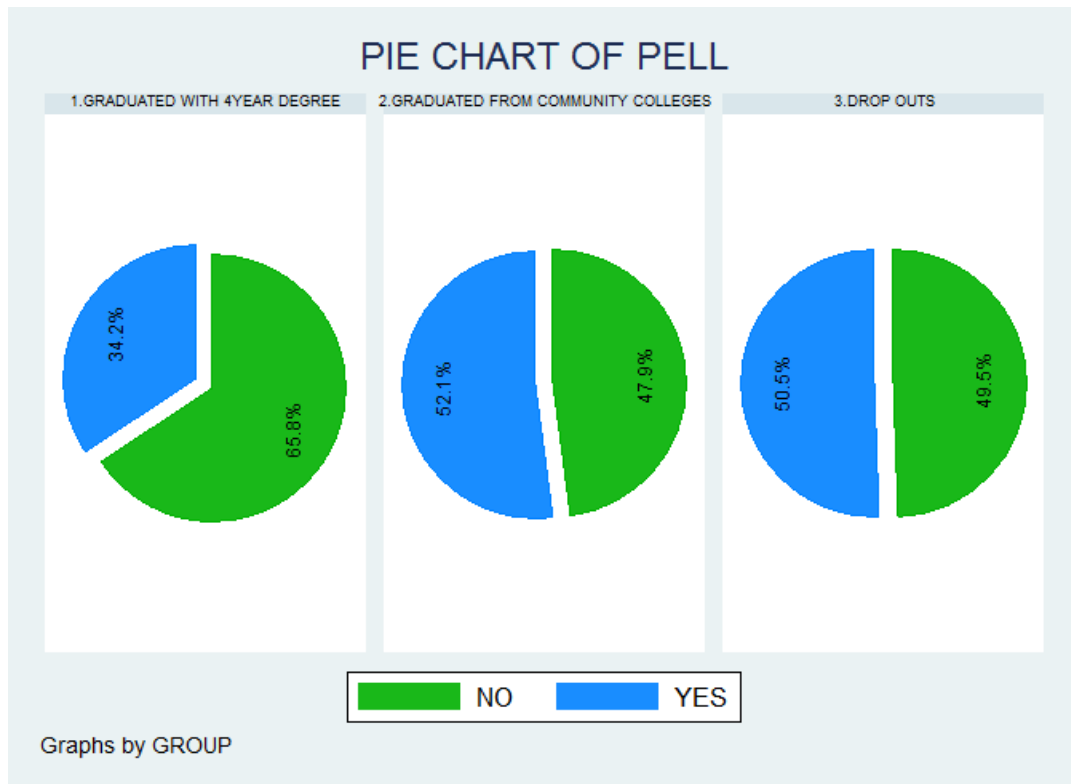
**Inference:** Across all the groups 36-48% of students selected Biology & pre professional as major. 22-29% of students selected Engineering as major. 19-23% of students selected Agsci as major. 2-3% of students selected chem & phys as major.



**FIGURE 6: PIE CHART OF DORM**

1) Among 4-year degree graduates: 75.3% lived in dorms and 24.7% did not live in dorms. 2) Among community college graduates: 79.2% lived in dorms and 20.8% of students did not live in dorms. 3) Among drop outs: 75% lived in dorms and 25% did not live in dorms.

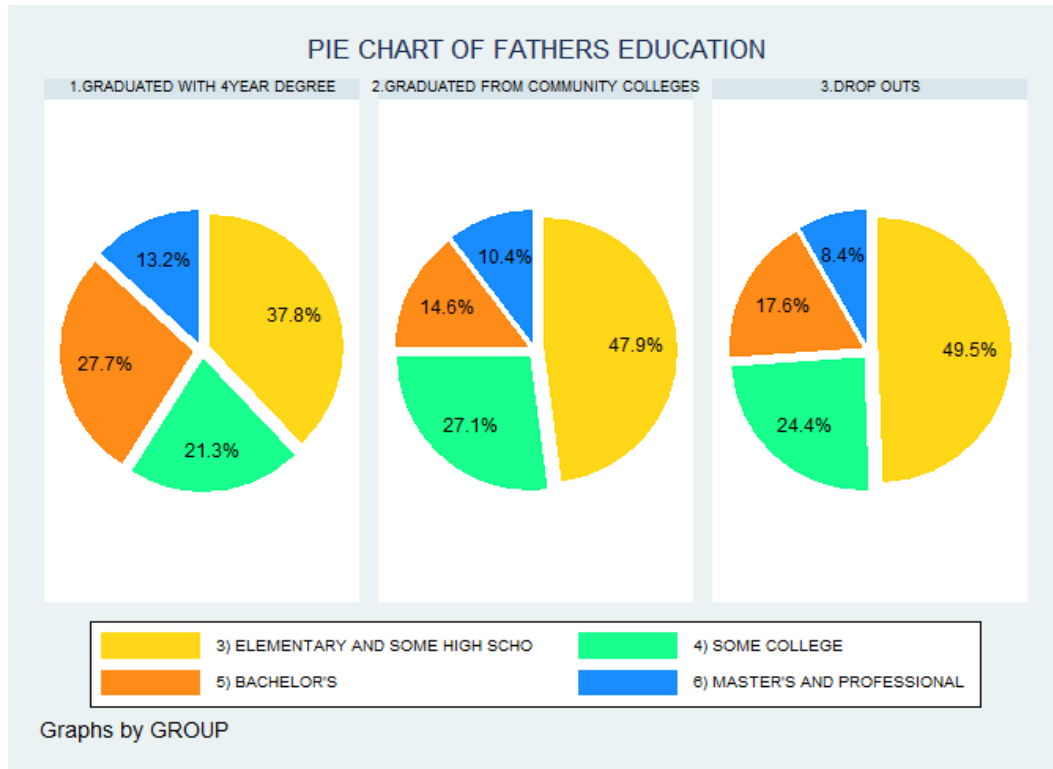
**Inference:** Living in or not living in dorms did not influence % graduation either with a 4-year degree or community college degree.



**FIGURE 7: PIE CHART OF PELL**

- 1) Among 4-year graduates: 65.8% received PELL, and 34.2% not received.
- 2) Among community college graduates: 52.1% received PELL, and 47.9% not received PELL.
- 3) Among drop outs: 50.5% received PELL, and 49.5% not received PELL.

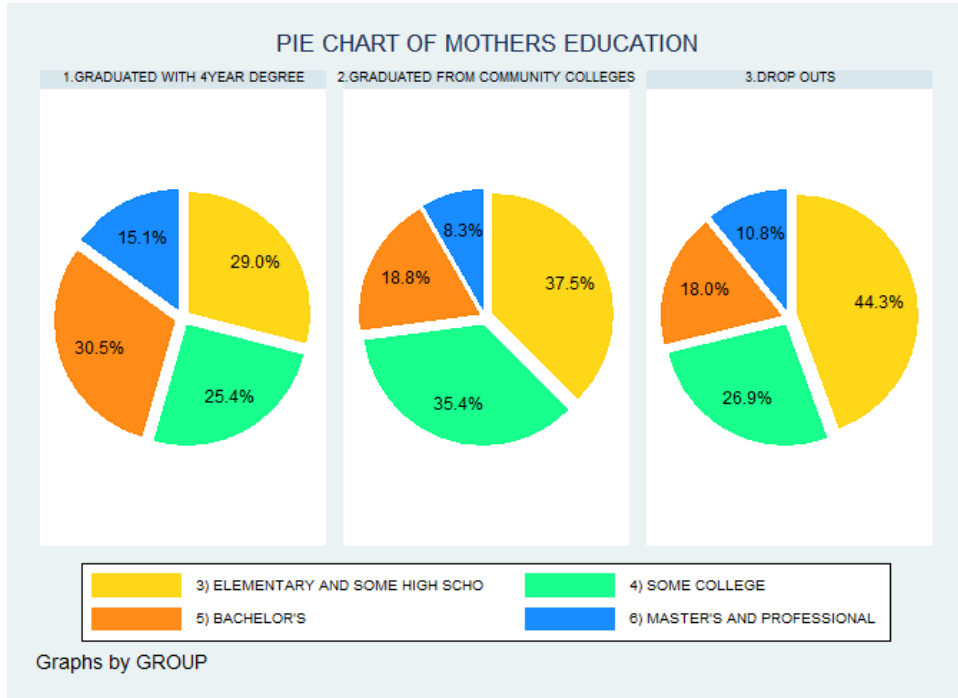
**Inference:** About  $\frac{1}{3}$  of 4-year graduates received PELL, while about  $\frac{1}{2}$  of the community college graduates and drop out were received PELL.



**FIGURE 8: PIE CHART OF FATHERS EDUCATION**

- 1) Among 4-year graduates: Fathers education of, 37.8% elementary and some high school, 27.7% bachelor's, 13.2% masters and professionals, 21.3% some college.
- 2) Community college graduates: Fathers education of 47.9% elementary and some high school, 14.6% bachelors, 10.4% masters and professionals, 27.1% some college.
- 3) Among Drop outs: Fathers education of 49.5% elementary and some high school, 17.6% bachelors, 8.4% masters and professionals, 24.4% some college.

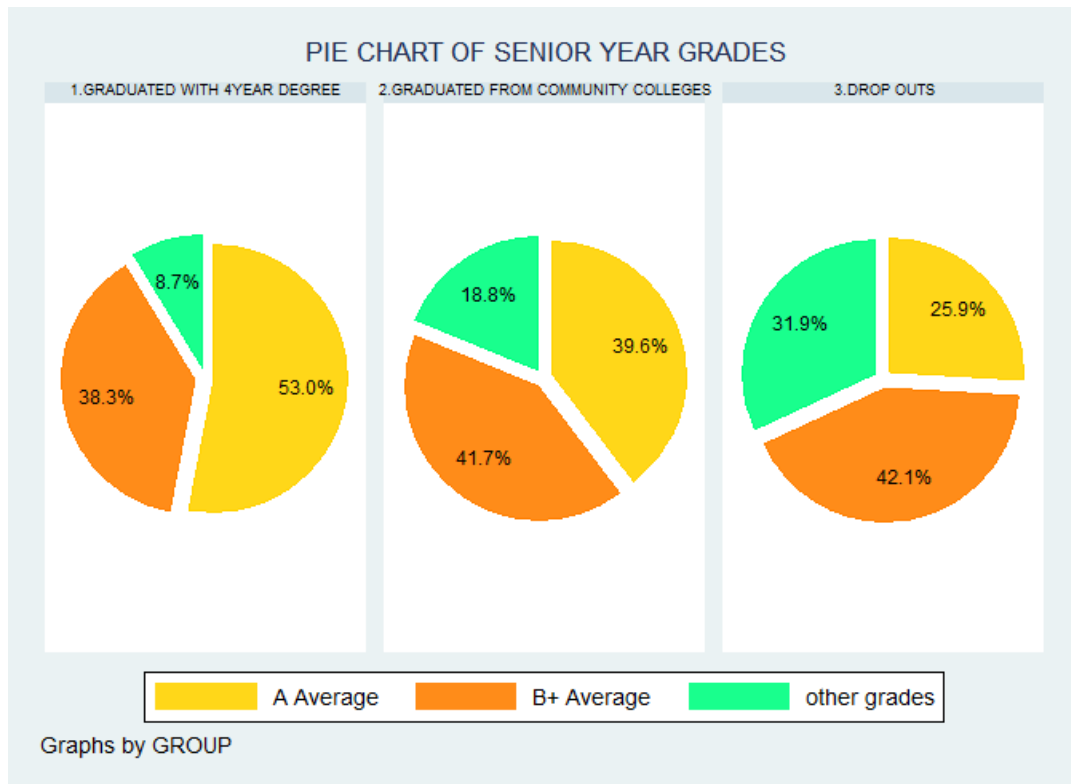
**Inference:** 10-12% of fathers have elementary education and some high school. 3-6% of fathers have some college education. 3-13% of fathers have bachelor's degree. 2-3% of fathers have a master's and professional degree.



**FIGURE 9: PIE CHART OF MOTHERS EDUCATION**

- 1) Among 4-year graduates: Mothers education of 29.0% elementary and some high school, 30.5% bachelor's, 15.1% masters and professionals, 25.4% some college.
- 2) Among Community college graduates: Mothers education of 37.5% elementary and some high school, 18.8% bachelors, 8.3% masters and professionals, 35.4% some college.
- 3) Among Drop outs: Mothers education of 44.3% elementary and some high school, 18% bachelors, 10.8% masters and professionals, 26.9% some college.

**Inference:** 8-15% of mothers have elementary education and some high school. 1-10% of mothers have some college education. 0.8-12% of mothers have bachelor's degree. 2-7% of mothers have a master's and professional degree.

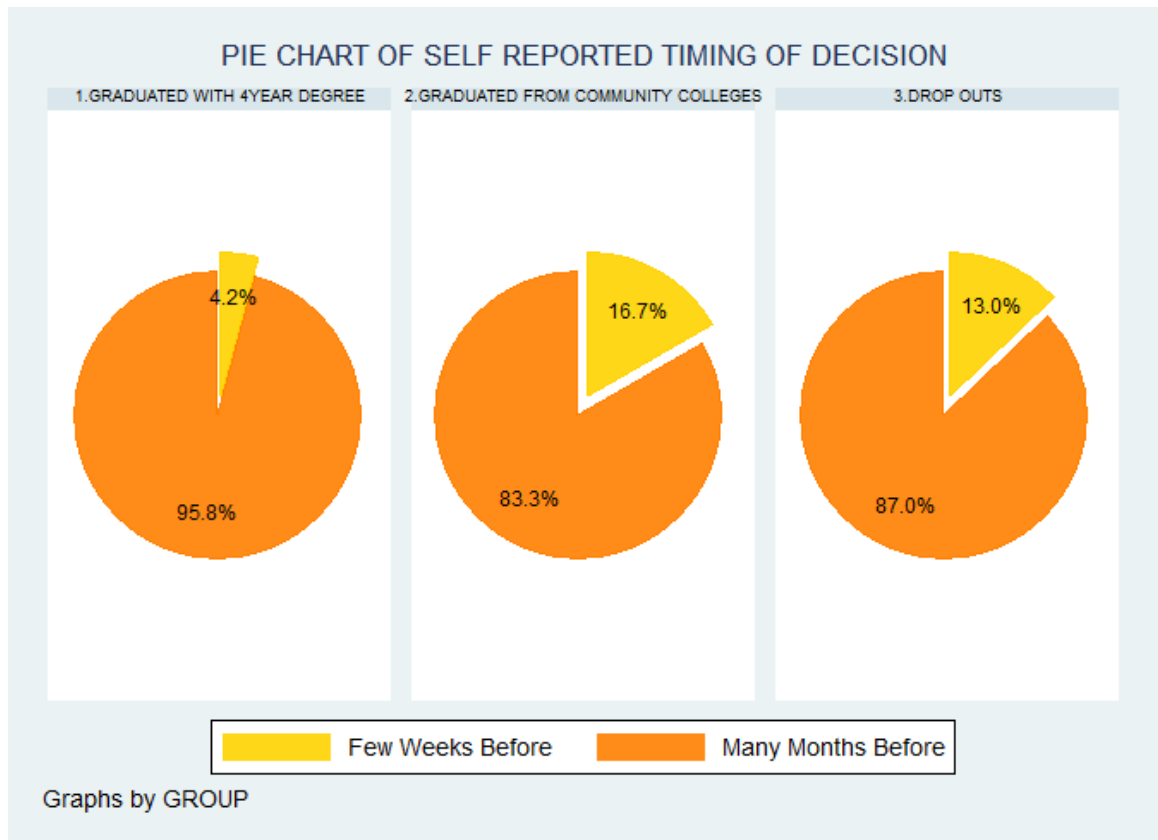


**FIGURE 10: PIE CHART OF SENIOR YEAR GRADES**

- 1) Among 4-year graduates: 53.0% had senior year grade A, 38.3% had grade B+ average, and 8.7% had other grade.
- 2) Community college graduates: 39.6% had senior year grade A, 41.7 % had grade B+ average and 18.8% had other grades.
- 3) 25.9% had grade A, 42.1 % had grade B+ average, and 31.9% had other grade.

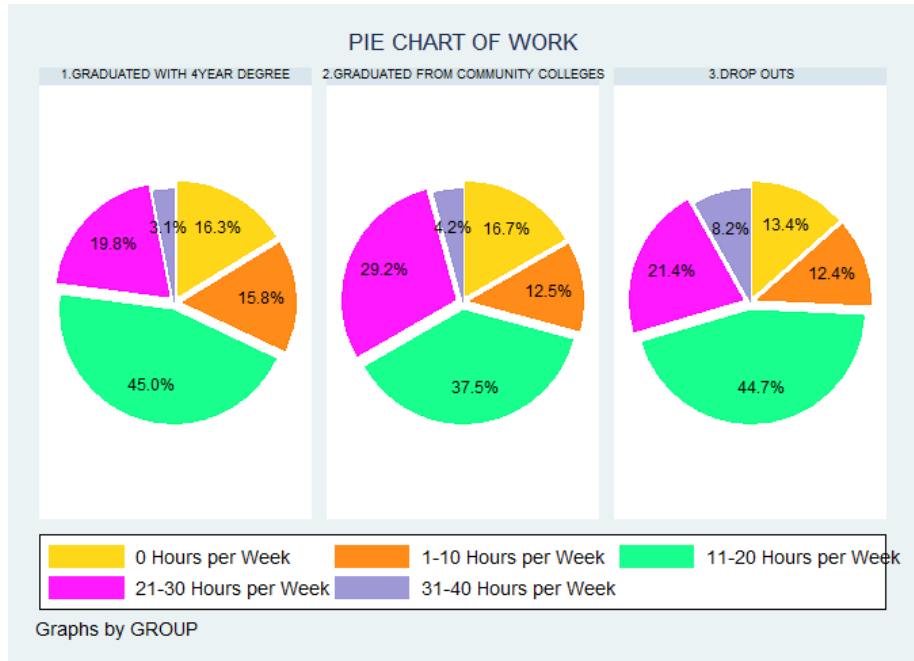
**Inference:** Proportion of students with C or lower grade in senior year is greater in drop outs (32%) than 4-year degree graduates (8.7%) or community college graduates (18.8%).





**FIGURE 11: PIE CHART OF SELF REPORTED TIMING OF DECISION**

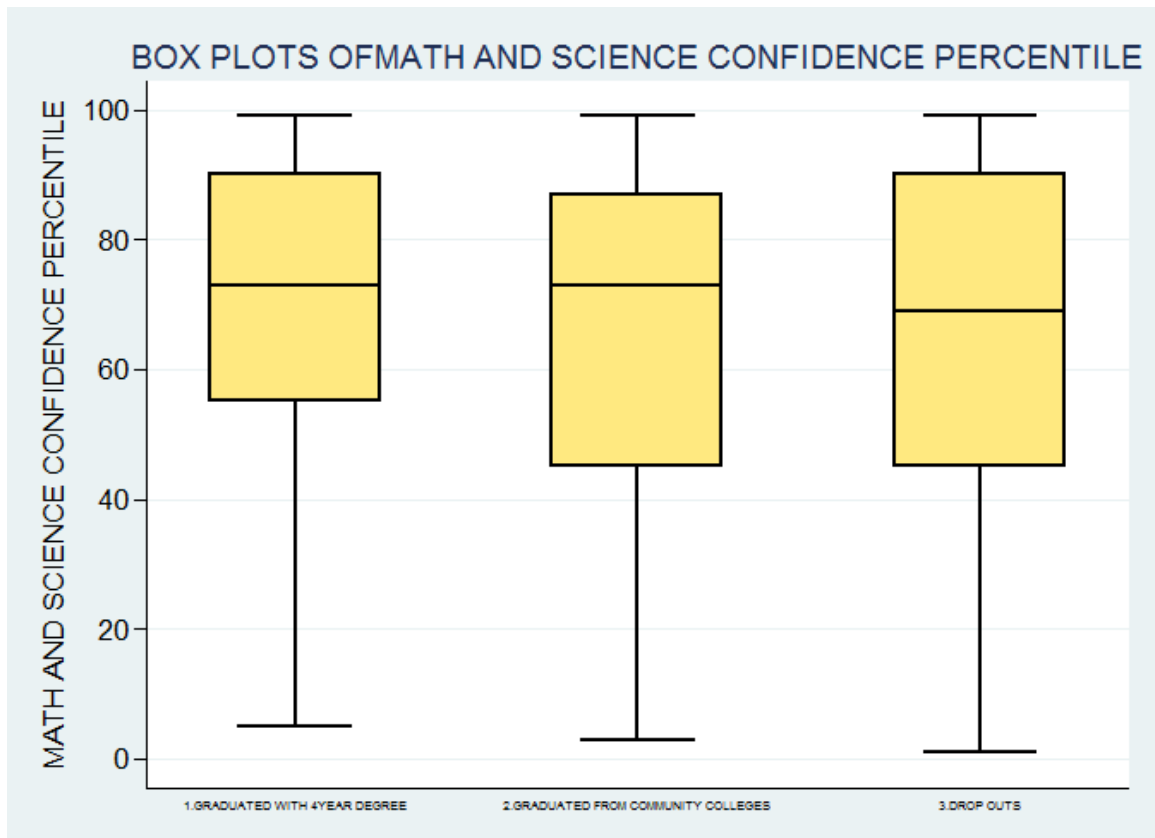
**Inference:** Greater percentage (95.8%) 4-year degree graduates have decided to join WTAMU than community college graduates (83.3%) and drop outs (87%).



**FIGURE 12: PIE CHART OF WORK**

1) Among 4-year graduates: 16.3% of 0 hours per week, 15.8% of 1-10 hours per week, 45.0% of 11-20 hours per week, 19.8% of 21-30 hours per week and 3.1% of 31-40 hours per week. 2) Among community college: 16.7% of 0 hours per week, 12.5% of 1-10 hours per week, 37.5% of 11-20 hours per week, 29.2% of 21-30 hours per week and 4.2% of 31-40 hours per week. 3) Among drop outs: 13.4% of 0 hours per week, 12.4% of 1-10 hours per week, 44.7% of 11-20 hours per week, 21.4% of 21-30 hours per week, 7.6% of 31-40 hours per week and 0.6% of 40+ hours per week.

**Inference:** Greater percentage (8.2%) drop outs worked 31-40 hours a week compared to community college graduates (4.2%), or 4-year degree graduates (3.1%).

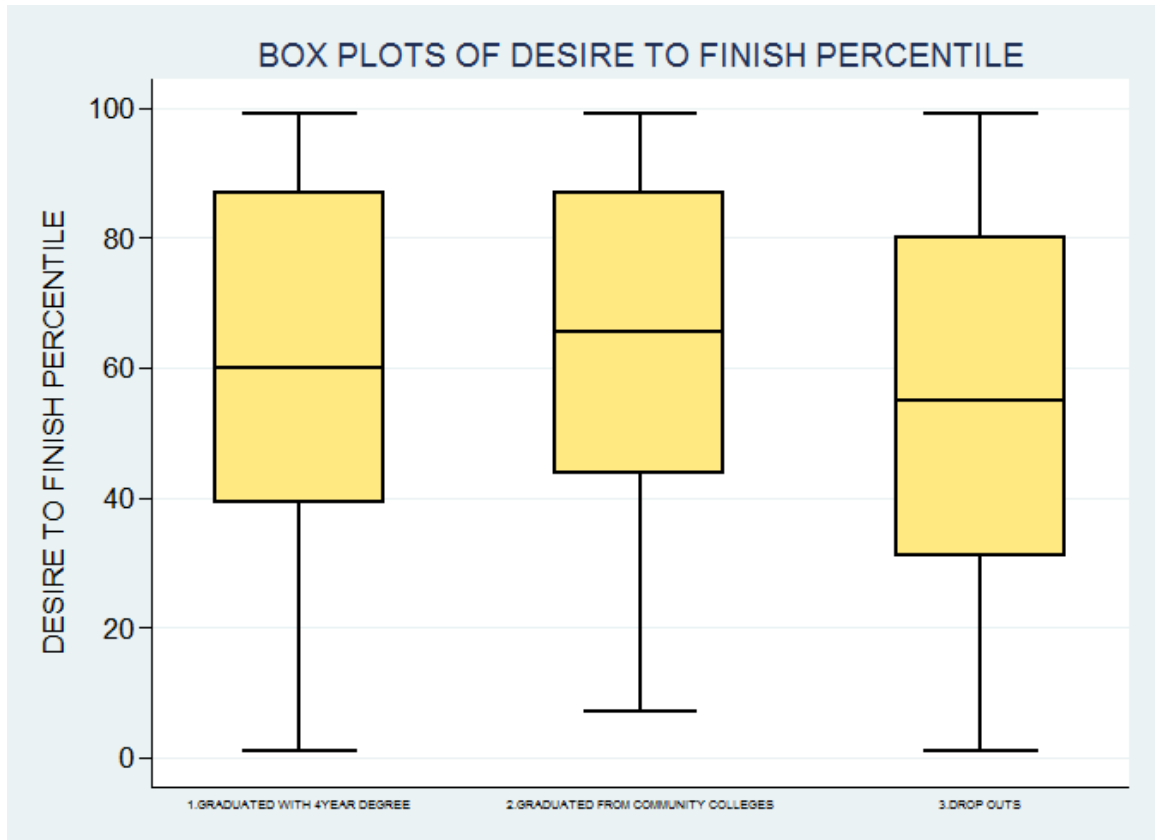


**FIGURE 13: BOX PLOTS OF MATH AND SCIENCE CONFIDENCE PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median math and science confidence percentile of 4-year graduates is 73 (90-55), of community college graduates is 73 (87-45), and of drop outs is 69 (90-45).

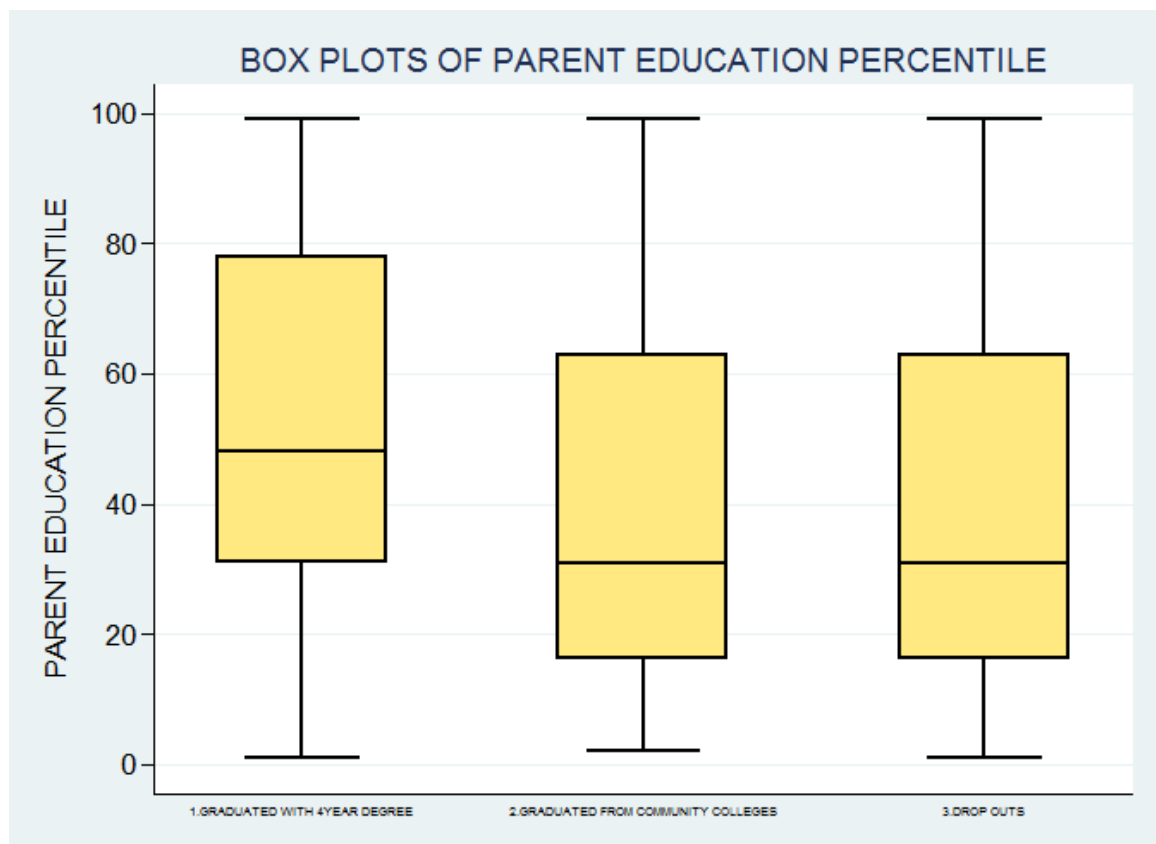
**Inference:** The median of math and science confidence for 4 year graduates and community college graduates is 4 percentile greater than drop outs.



**FIGURE 14: BOX PLOTS OF DESIRE TO FINISH PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). Median desire to finish percentile of students was 60 (87 - 39), 65 (87 - 44) and 55 (80-31) for 4-year degree graduates, community college graduates and drop out students.

**Inference:** Median desire to finish percentile of drop out students was 5 -10 percentile less than 4-year degree graduates or community college graduates.

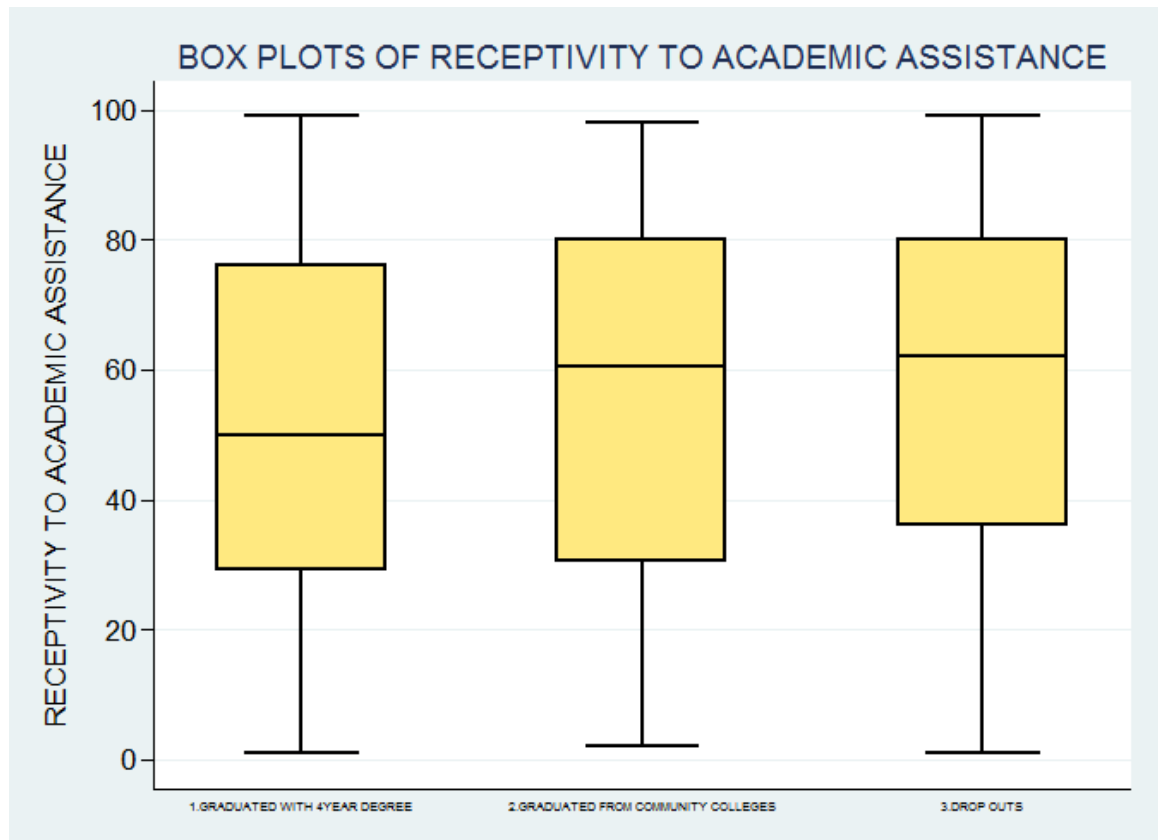


**FIGURE 15: BOX PLOTS OF PARENT EDUCATION PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median parent education percentile of 4-year graduates is 48 (78-31), of college graduates is 31(63-16), of drop outs is 31 (63-16).

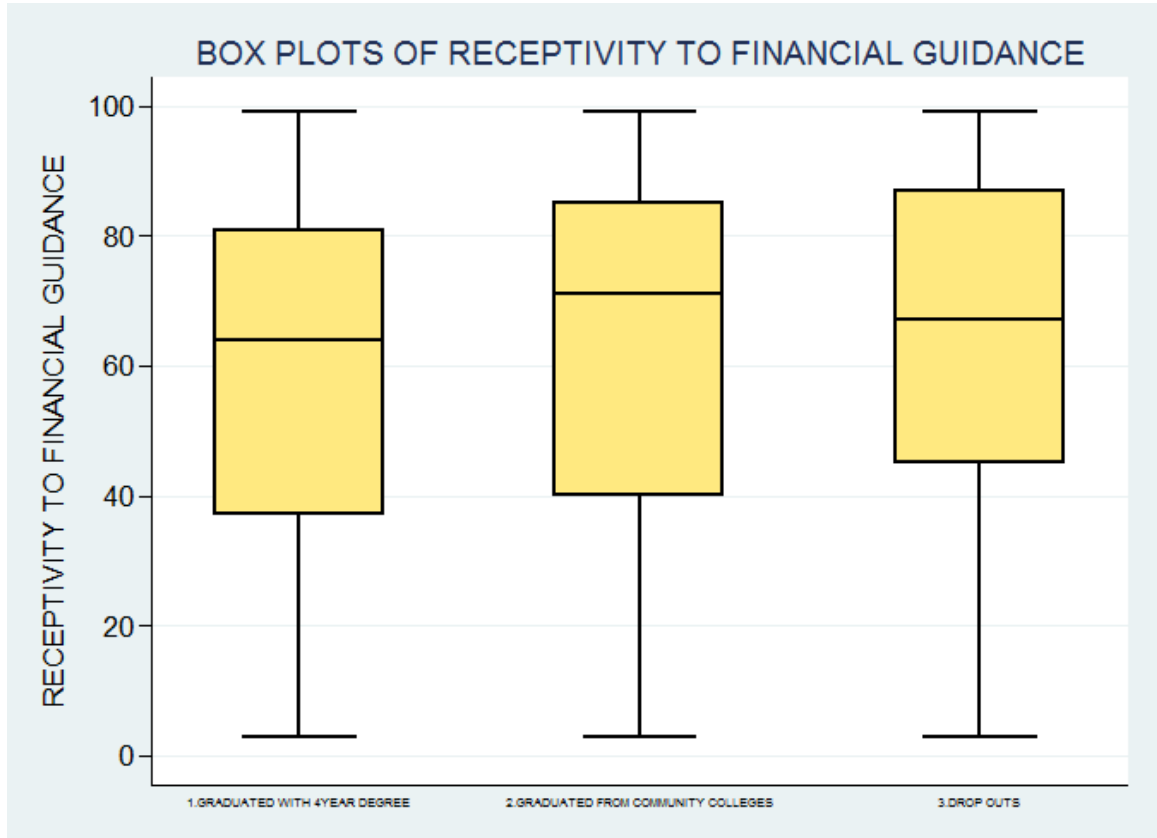
**Inference:** Median education of undergraduate level student parent's is 17 percentile greater than parents of community college graduates or drop outs.



**FIGURE 16: BOX PLOT OF RECEPTIVITY TO ACADEMIC ASSISTANCE**

Data are median with the third quartile (Q3) and first quartile (Q1). Median receptivity to academic assistance of students was 50 (76 - 29), 60.5(80 – 30.5), and 62(80- 36) for 4-year degree graduates, community college graduates and drop out students.

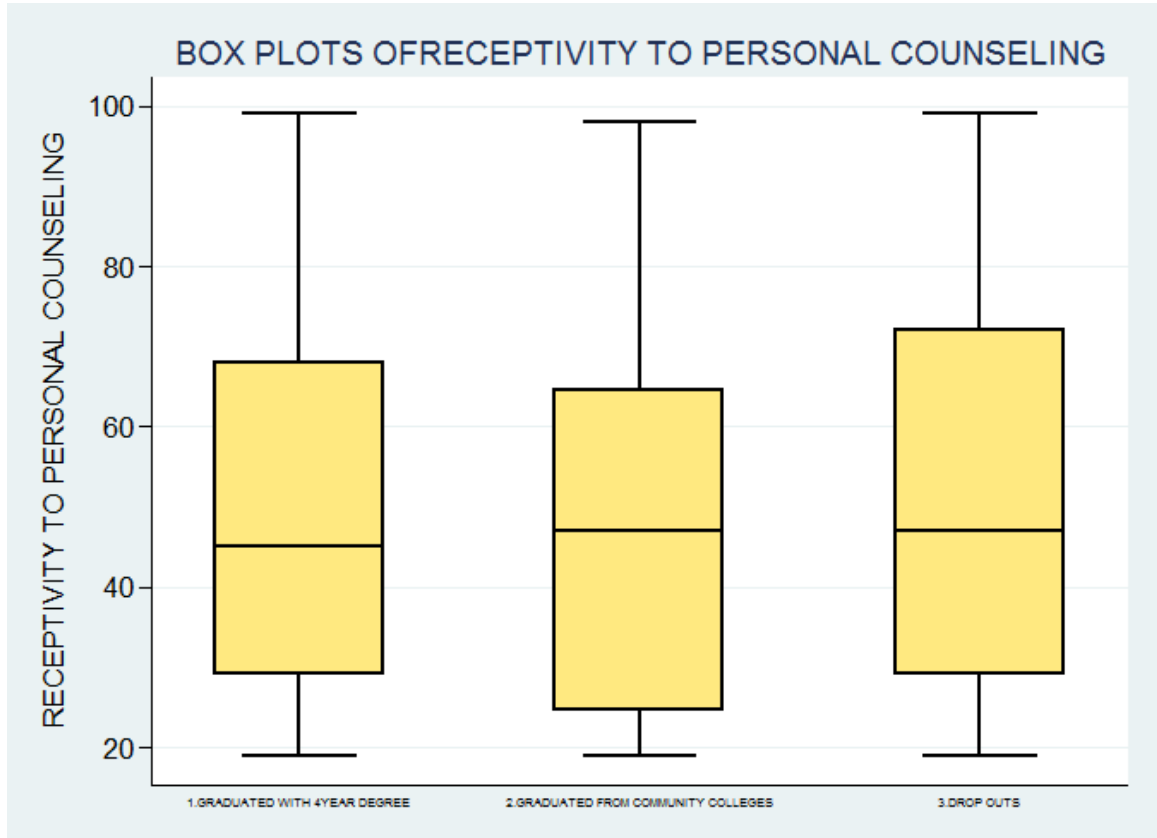
**Inference:** Median receptivity to academic assistance community college graduates or drop out students was at least 10 - 12 percentile higher than the 4-year degree graduates



**FIGURE 17: BOX PLOTS OF RECEPTIVITY TO FINANCIAL GUIDANCE**

Data are median with the third quartile (Q3) and first quartile (Q1). Median 64 (81-37) for all 4 year graduates. 72 (85-40) for all community graduates and 67 (87-45) for drop outs.

**Inference:** Drop outs median receptivity to financial guidance was at least 2 – 3 percentile less than graduates with 4-year degree students or community college.

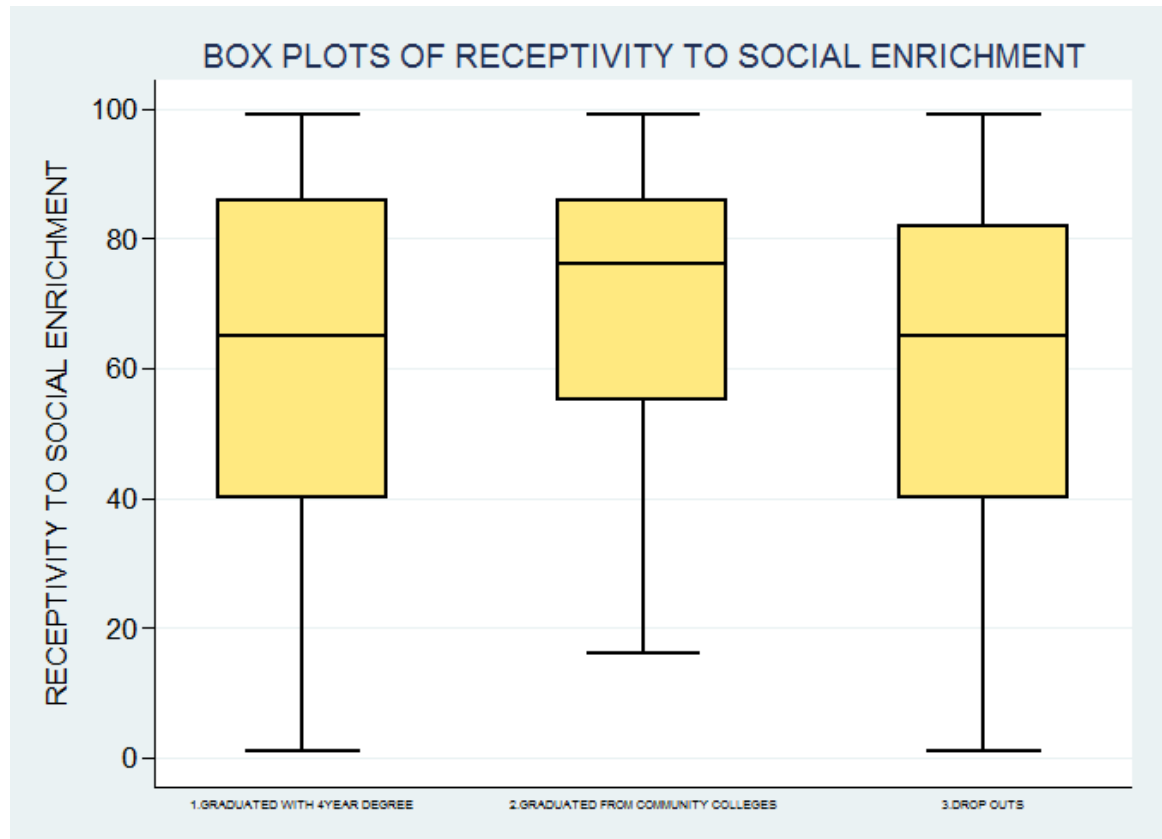


**FIGURE 18: BOX PLOTS OF RECEPTIVITY TO PERSONAL COUNSELING**

Data are median with the third quartile (Q3) and first quartile (Q1). Median of receptivity to personal counseling is 45 (68-29) for 4 year graduates, 47 (64.5-24.5) for community graduates and 47 (72-29) for drop outs.

**Inference:** Graduated with 4-year degree students' median Receptivity to personal counseling was at least 1 - 4 percentile less than community college and drop outs.

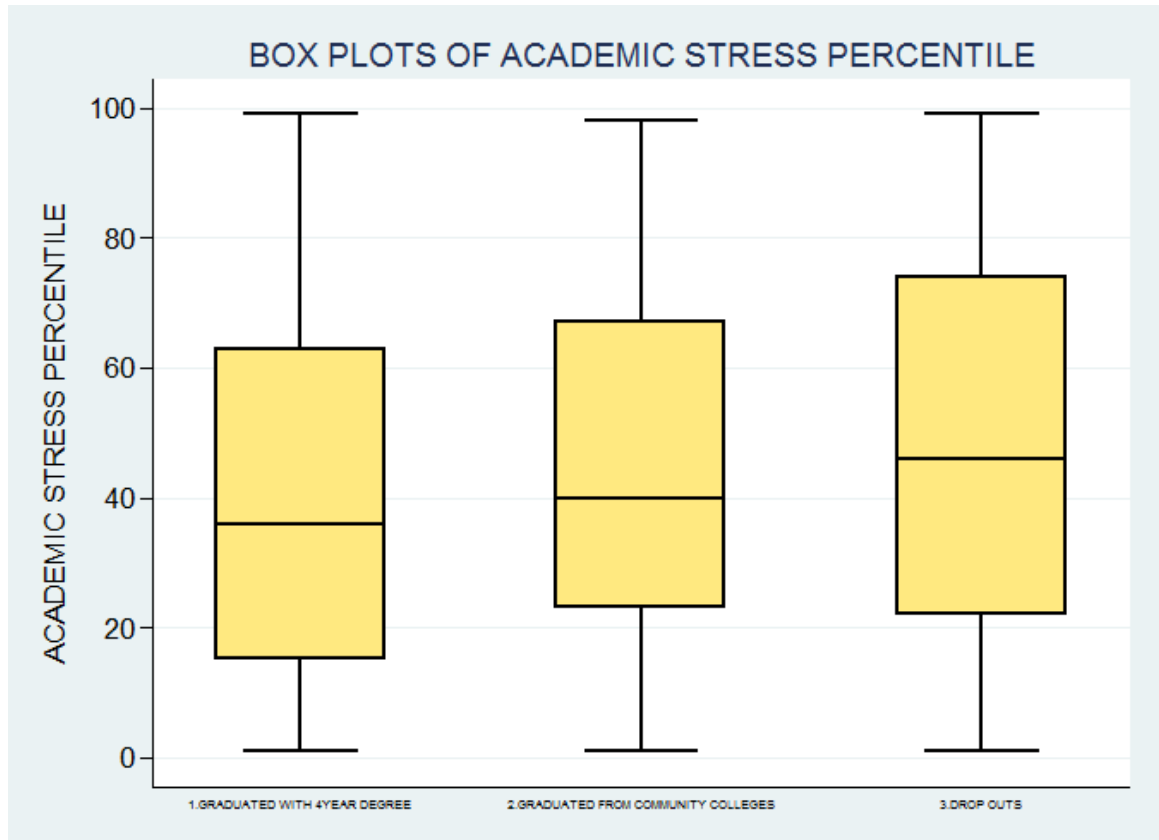




**FIGURE 19: BOX PLOT OF RECEPTIVITY TO SOCIAL ENRICHMENT**

Data are median with the third quartile (Q3) and first quartile (Q1). 1) The median for all 4 year graduates is 65, the median for all community graduates is 76 and the median for all dropped outs from universities and community colleges is 65. 2). The interquartile range for group 1 is 46(86-40); for group 2 is 31(86-55); for group 3 is 32(82-40).

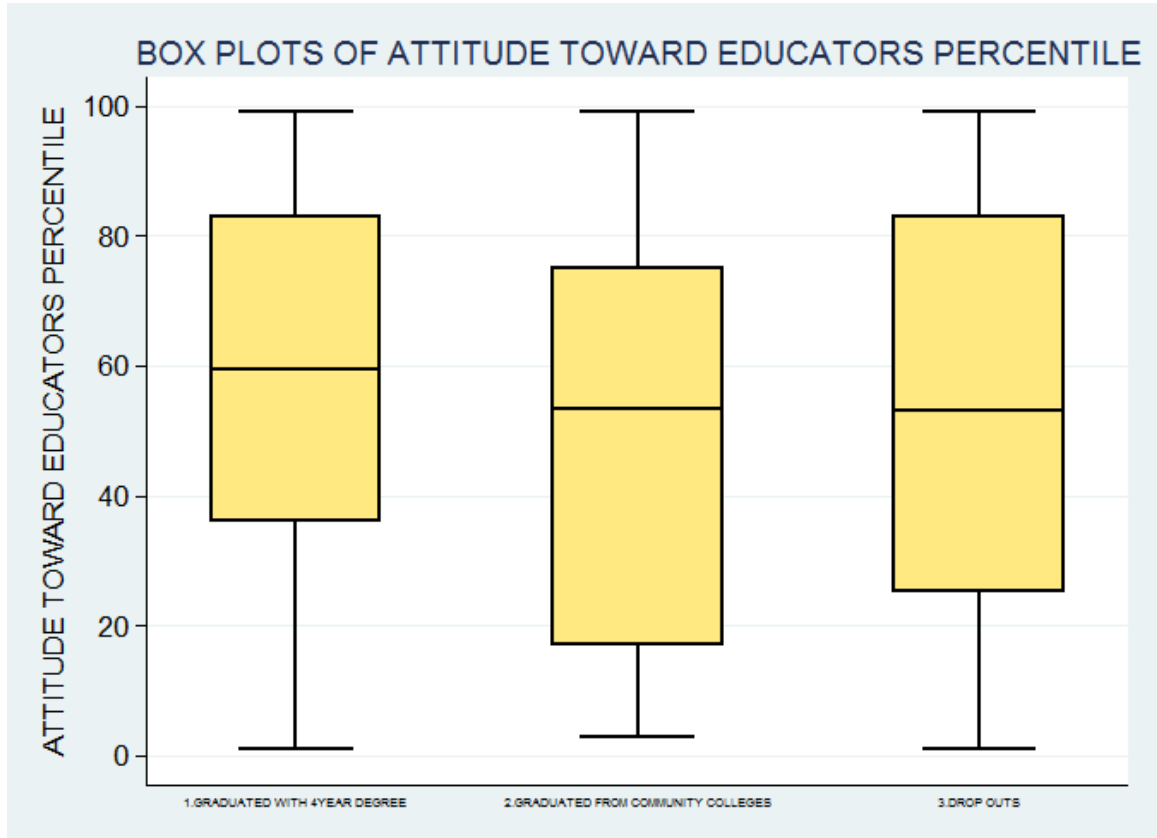
**Inference:** Community college students' median Academic stress percentile was at least 1- 15 percentile less than Graduated with 4-year degree students and drop outs.



**FIGURE 20: BOX PLOTS OF ACADEMIC STRESS PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). The median of academic stress percentile is 36 (63-15) for 4-year graduates, 36 (67-23) for community graduates, and 40 (74-22) for drop outs.

**Inference:** Community college students' median academic stress percentile is at least 4 - 8 percentile less than 4-year degree graduates or drop out students.

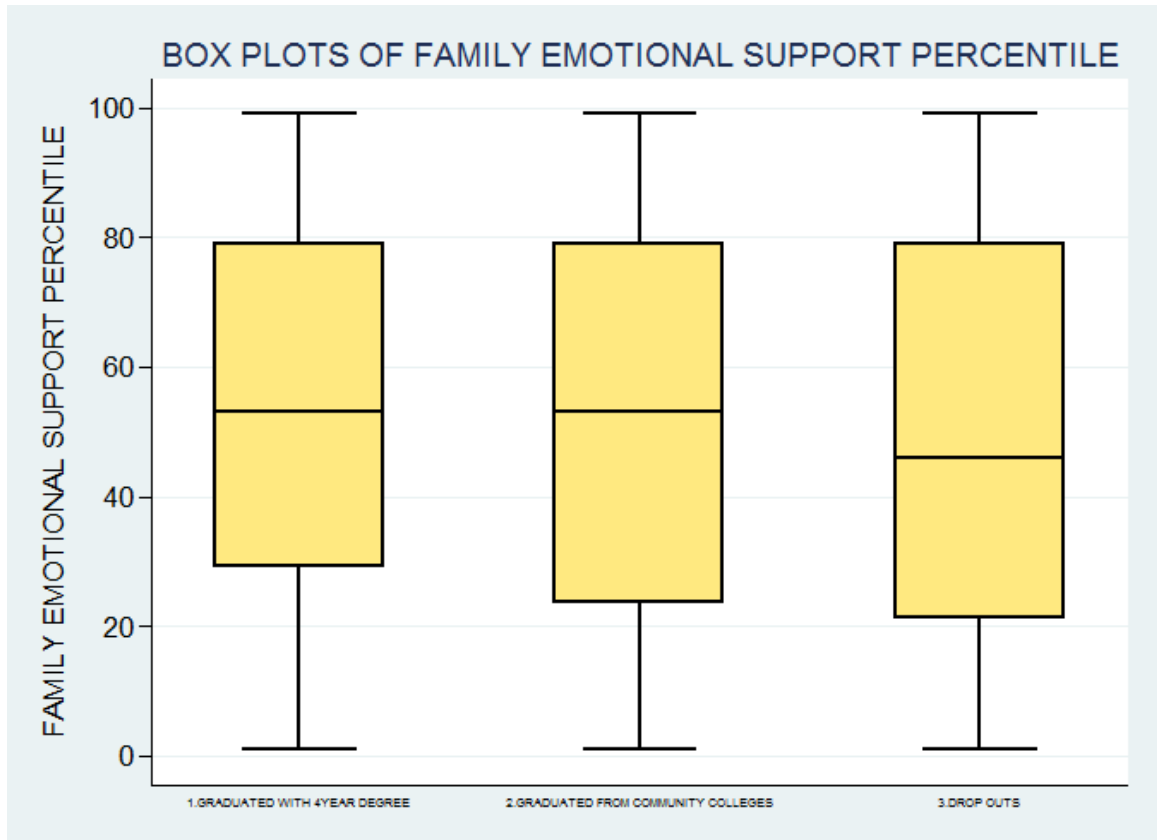


**FIGURE 21: BOX PLOTS ATTITUDE TOWARD EDUCATORS PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median attitude toward educators percentile is 59.5 (83-36) 4-year graduates, 53 (75-17) for community college graduates, and 53 (83-25) for drop outs.

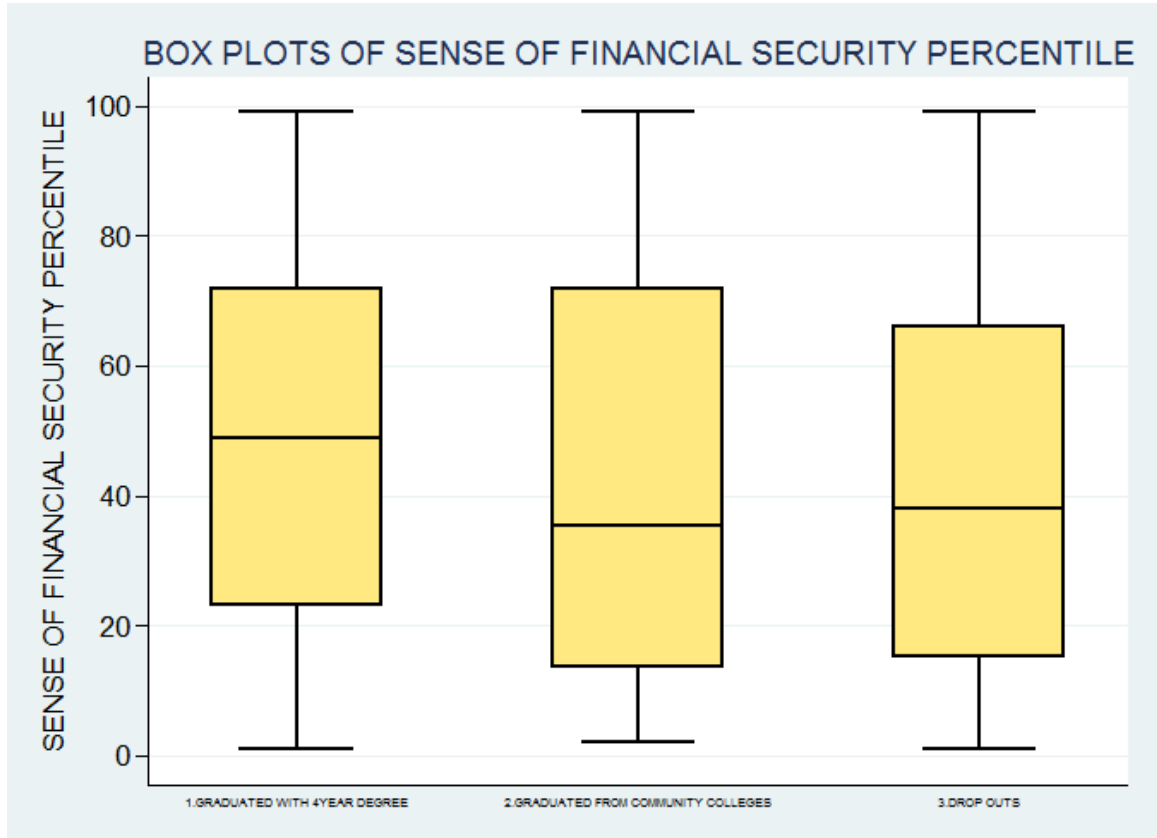
**Inference:** Graduated with 4-year degree students' median Attitude toward educators percentile was at least 11 percentile less than community college or drop outs.



**FIGURE 22: BOX PLOTS OF FAMILY EMOTIONAL SUPPORT PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). Median family emotional support percentile is 53 (79 - 29), 52 (79 - 24) and 43 (79-21) for 4-year degree graduates, community college graduates and drop out students, respectively.

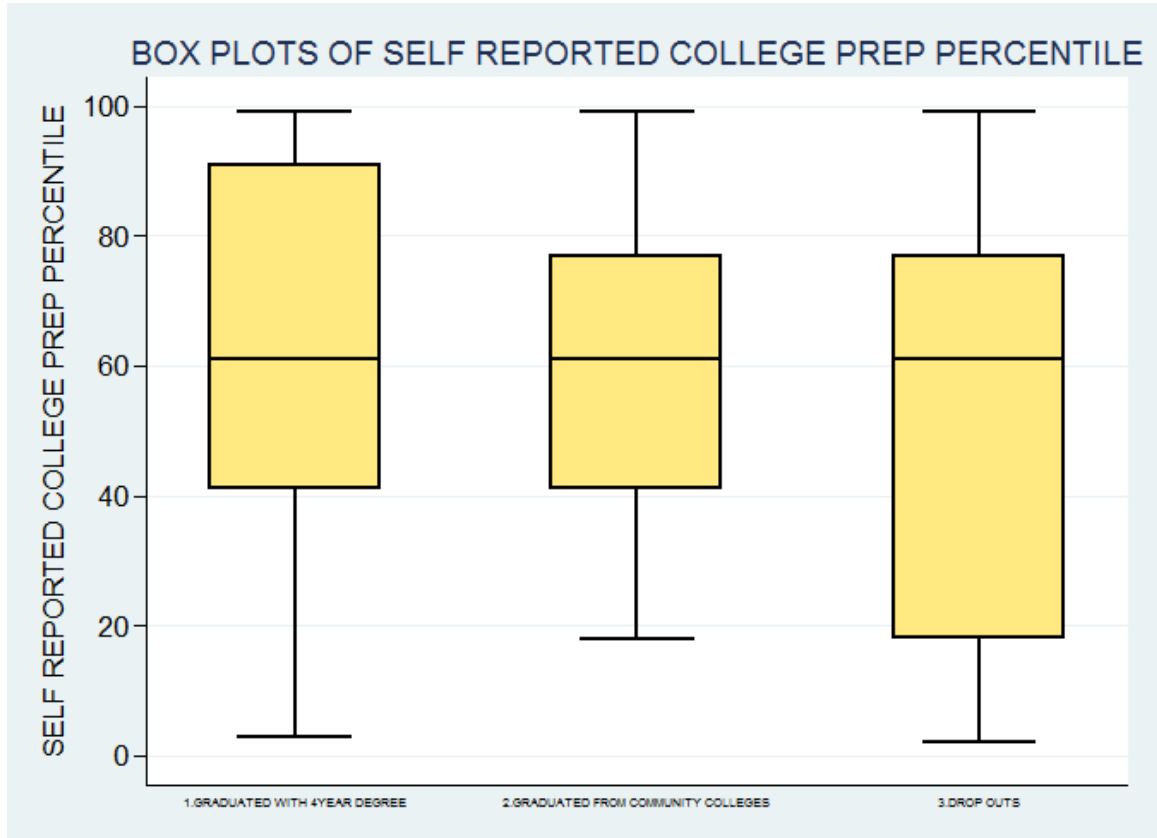
**Inference:** Drop out students median family emotional support percentile was at least 9 percentile less than graduated students with 4-year degree or community college.



**FIGURE 23: BOX PLOTS OF SENSE OF FINANCIAL SECURITY PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). The median financial security percentile is 49 (72-23) for 4-year degree graduates, 39 (72-13.5) for community college graduates, and 38 (66-15) for drop outs.

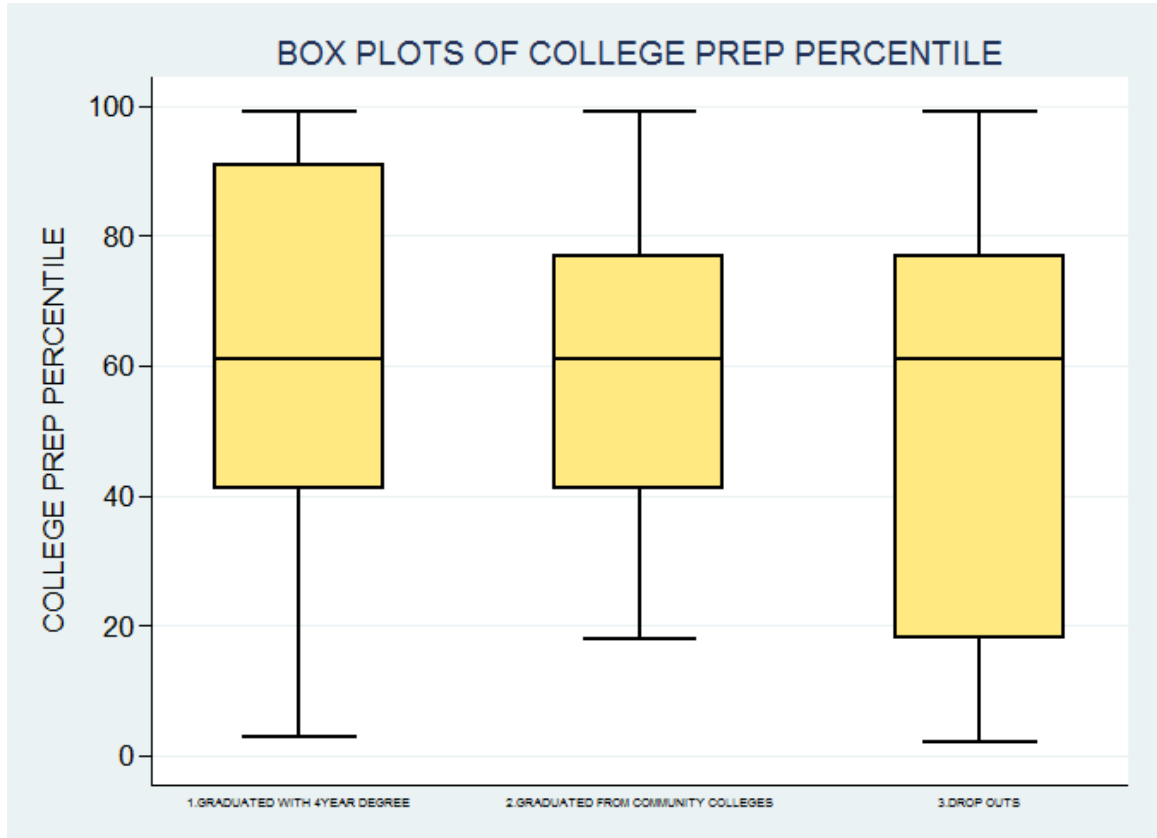
**Inference:** 4-year degree graduates median sense of financial security percentile was at least 11-12 percentile greater than community college graduates or drop outs.



**FIGURE 24: BOX PLOTS OF SELF REPORTED COLLEGE PREP PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). Self-reported college prep percentile median for 4- year graduates is 61 (91-41), for community graduates is 61 (77-41) , and for drop outs is 61 (77-18).

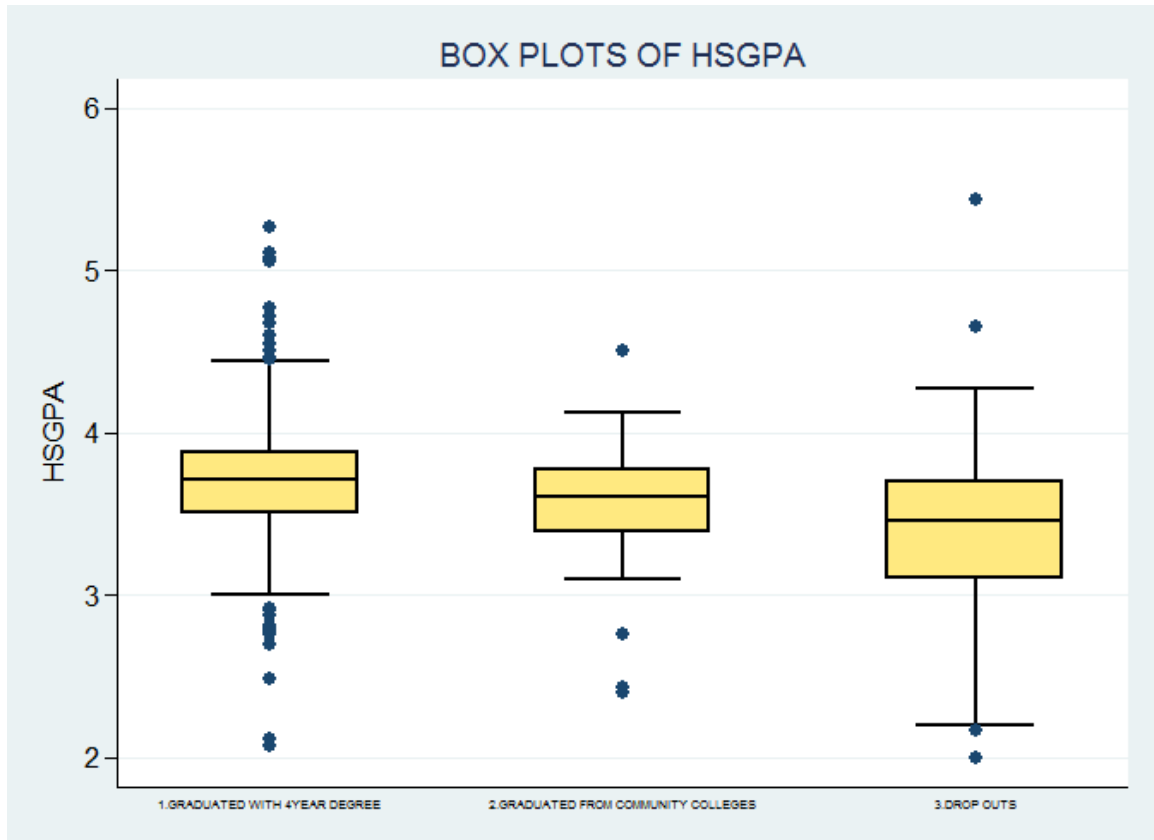
**Inference:** Self-reported college prep percentile do not affect graduation or drop out.



**FIGURE 25: BOX PLOTS OF COLLEGE PREP PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). The median college prep percentile for 4-year graduates is 61 (91-41), community college graduates is 61 (77-41), and drop outs is 61 (77-18).

**Inference:** College prep percentile of 4-year graduates, community college graduates and drop outs is similar.

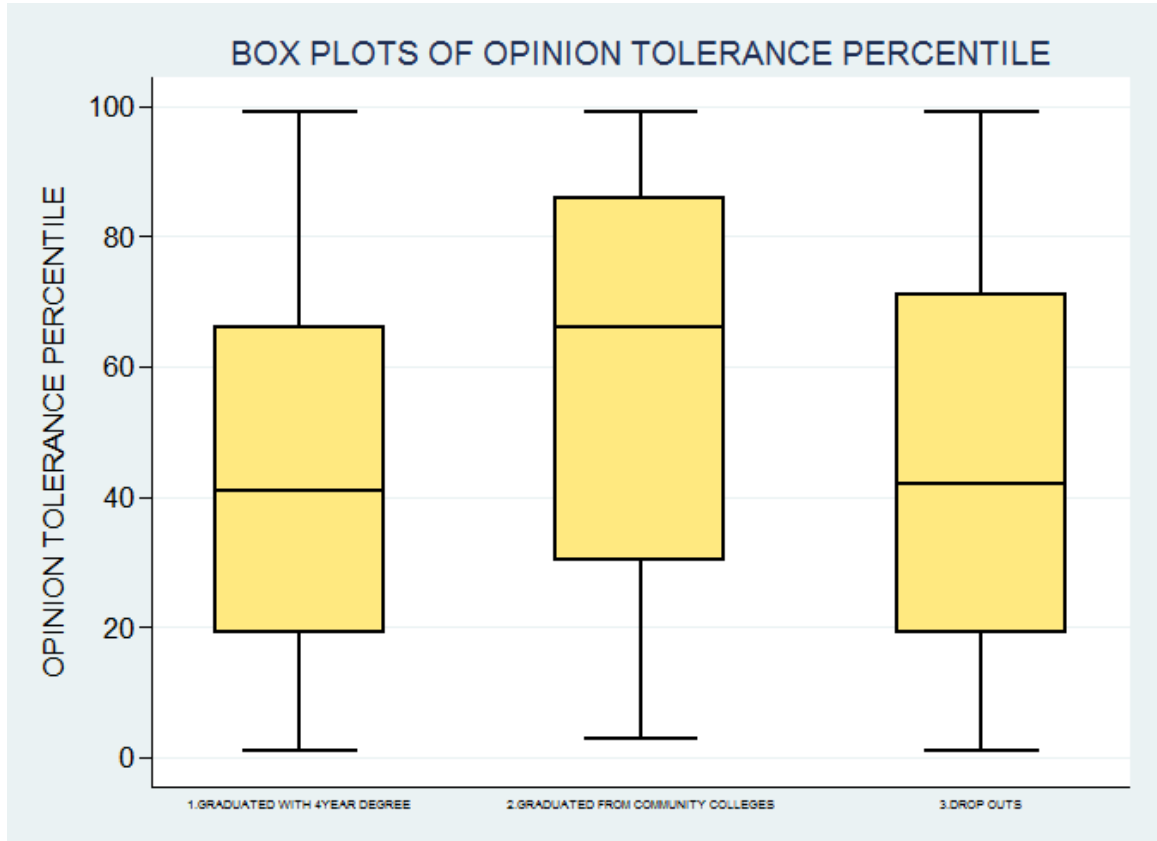


**FIGURE 26: BOX PLOTS OF HSGPA**

Data are median with the third quartile (Q3) and first quartile (Q1). Median high school grades percentile of students was 3.715, 3.605 and 3.46 for 4-year degree graduates, community college graduates and drop out students.

**Inference:** Descriptive analysis of data demonstrated that median GPA percentile of drop out students is 3-7 percentile less than the GPA percentile of 4-year degree graduates and community college graduates.

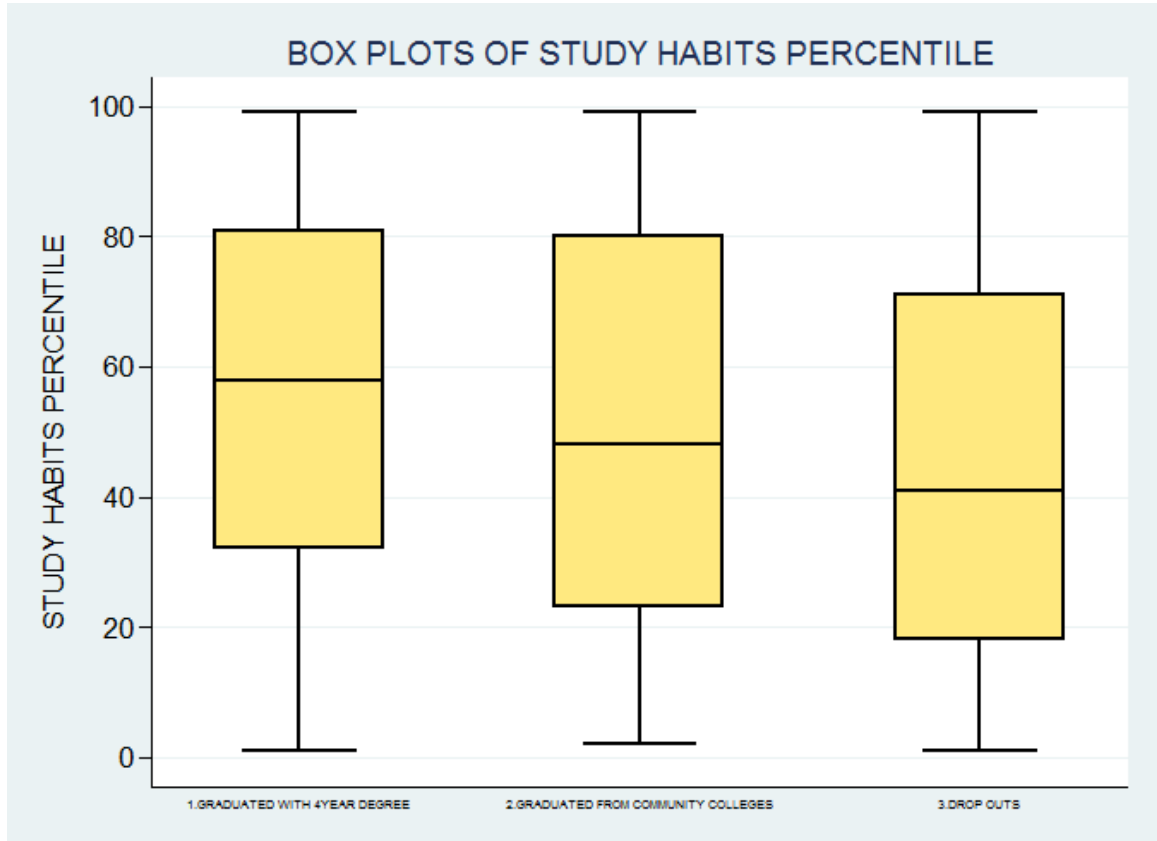




**FIGURE 27: BOX PLOTS OF OPINION TOLERANCE PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). The median opinion tolerance percentile for 4-year graduates is 41 (66-19), community college graduates is 66 (86-30) and, drop outs is 42 (71-19).

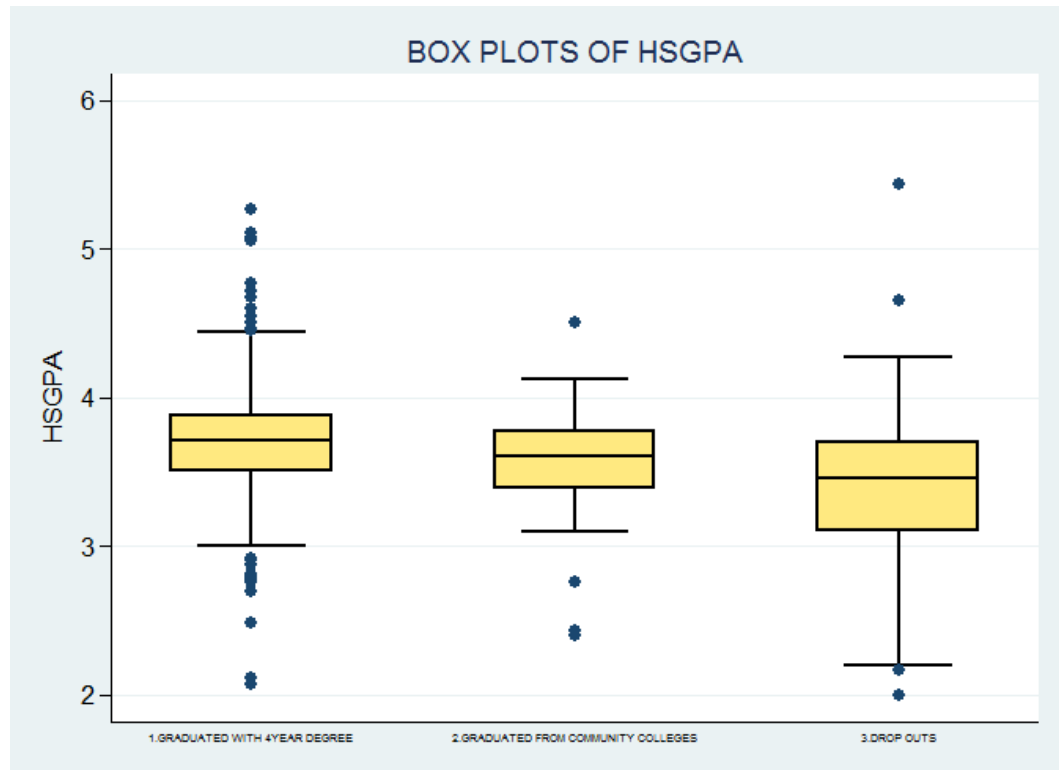
**Inference:** Community college graduates median opinion tolerance percentile was at least 15-16 percentile less than 4-year degree graduates drop outs.



**FIGURE 28: BOX PLOTS OF STUDY HABITS PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1). The median habit percentile for 4-year graduates is 58 (81-32), community graduates is 48 (80-23) and the median drop outs is 41 (71-18).

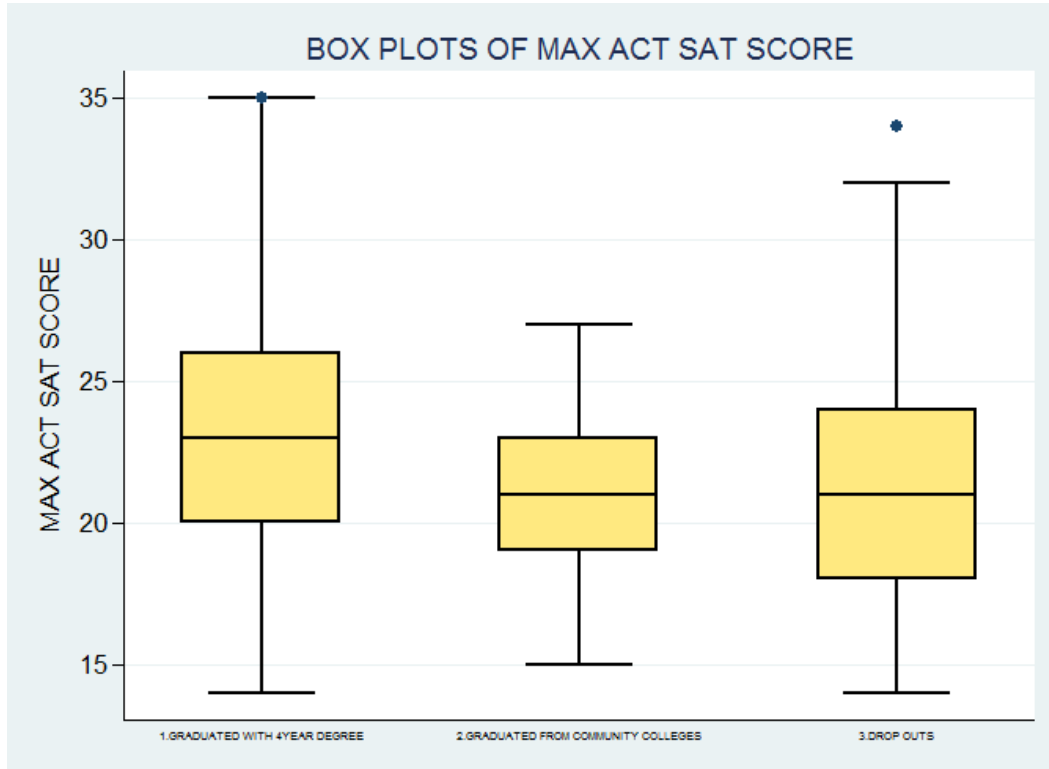
**Inference:** 4-year degree students' median study habits percentile was at least 4-8 percentile greater than community college graduates or drop outs



**FIGURE 29: BOX PLOTS OF HIGH SCHOOL GPA**

Data are median with the third quartile (Q3) and first quartile (Q1). 1) Median high school GPA of graduates is 3.72 (3.88 - 3.5), 2) median high school GPA of community college graduates is 3.6 (3.78 - 3.39), and 3) median high school GPA of drop outs is 3 (3.7 - 3.1).

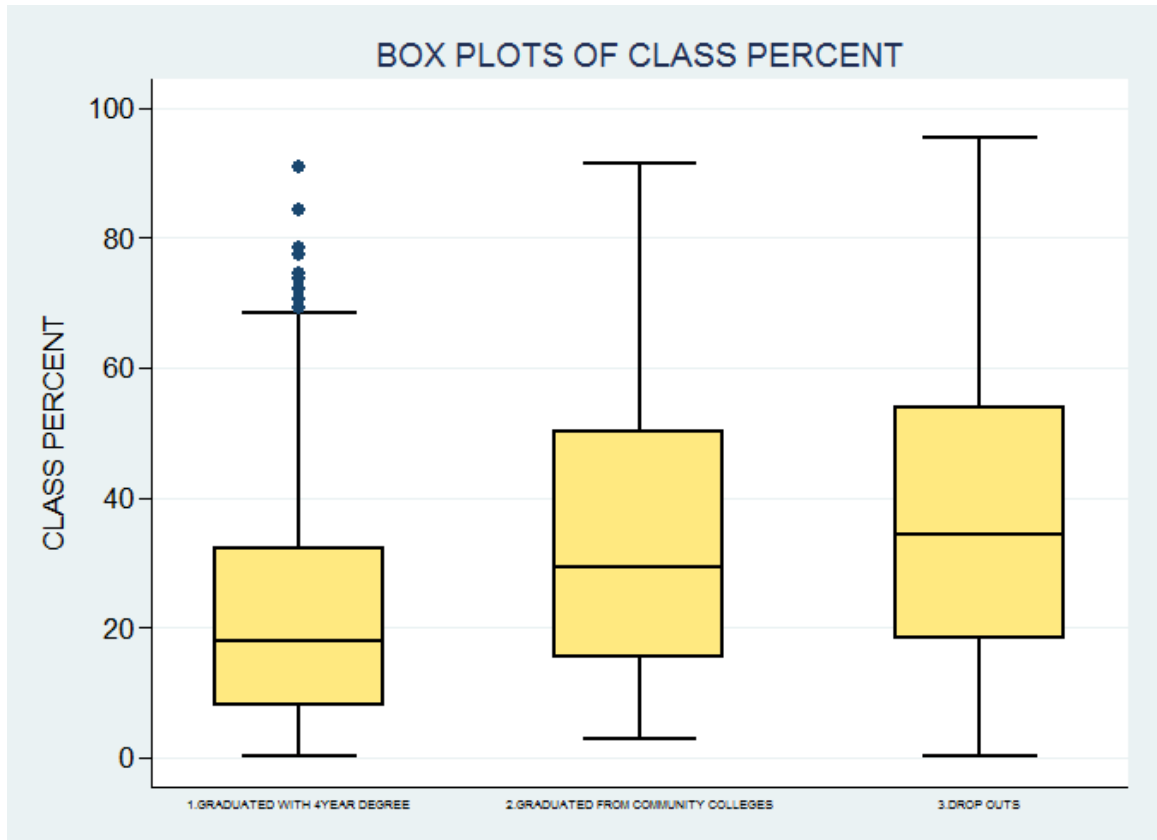
**Inference:** The median HSGPA of 4-year degree graduates was at least 0.6 - 0.72 percentile greater than the community college graduates or drop outs



**FIGURE 30: BOX PLOTS OF MAX ACT SAT SCORE**

Data are median with the third quartile (Q3) and first quartile (Q1). Median ACT or SAT score of students was 23 (26 - 20), 21 (23 - 19) and 21 (24-18) for 4-year degree graduates, community college graduates and drop out students, respectively.

**Inference:** Median Max ACT or SAT score percentile of 4-year degree graduates was 2 percentile greater than community college graduates or drop outs.



**FIGURE 31: BOX PLOTS OF CLASS PERCENT**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median Class percent is 17.9 for 4-year degree graduates, 29.3 for community graduates, and 34.2 for drop outs.

**Inference:** The median class percent of 4-year degree graduates students was at least 11-16 percentile less than community college graduates or drop outs.

### Development of Retention to Graduation Model

The first logistic regression model developed investigates factors affecting students' retention in higher education. The second model investigates factors impacts retention to graduation in a STEM degree program. Both logistic regression models are constructed using the LOGIT procedure in Stata. A threshold of 0.25 ( $\alpha = 0.25$ ) is used to identify independent variables in a single predictor logistic model with significant influence to be included in the preliminary complete model. Then with all significant independent variables, logistic regression is performed with a backwards elimination process to identify and remove the least significant independent variable, and this process was repeated until there was no non-significant independent variable. The final regression model is constructed using all significant predictors. A threshold of 0.05 ( $\alpha = 0.05$ ) was used to identify all independent variables with significant association.

### **Graduation in higher education**

**TABLE 1: SINGLE LOGISTIC REGRESSION ANALYSIS OUTPUT BETWEEN UNDERGRADUATE STUDENT GRADUATION AND INDEPENDENT VARIABLES**

<b>VARIABLE</b>	<b>COEFFICIENT</b>	<b>STD.ERROR</b>	<b>PVALUE</b>
Gender	.4979306	.1284064	0.001
Race 2	-.6660138	.1564378	0.000

3	-.9872185	.2704121	0.000
4	-.3458964	.2557721	0.176
Transfer Percentile	-.0042512	.0027348	0.120
Desire Percentile	.0061602	.0021842	0.005
Academic Assistan	-.0056389	.0022865	0.014
Financial Guidance	-.00499	.0023237	0.032
Personal Council	-.0036771	.0025664	0.152
Social Enrichhment	.0032365	.0023622	0.171
Stress percentile	-.0093283	.0022058	0.000
Educator percentile	.0050996	.0021763	0.019
Family Emotional percentile	.0066612	.0020391	0.001
Financial Security Percentile	.0077232	.0021855	0.000
Self Report College Prep Percentile	.0170813	.0024602	0.000
Math and Science Confidence Percentile	.0041778	.0024599	0.089

College Prep Percentile	.0172393	.0024664	0.000
High School Grade Percentile	.0314936	.0032827	0.000
Intellectual Interest Percentile	.0037408	.0021036	0.075
Study Habits Percentile	.0123835	.0021482	0.000
Desire to Transfer	-.0045837	.0027468	0.095
Max ACTSAT score	.1251162	.0176723	0.000
HSGPA	1.874639	.1941435	0.000
Class Percent	-.0307871	.0033695	0.000
PELL	-.5976081	.1290076	0.000
Major Code			
2	-.9470159	.4096071	0.021
3	-.6369769	.3407131	0.062
4	-.20319	.3326289	0.541
5	-.2707325	.4988405	0.587
6	-.2125725	.345602	0.539



Fathers Education	.2657199	.0619583	0.000
Mothers Education	.3087194	.0619016	0.000
Transfer Desire	-.2352746	.1312698	0.073
Senior Year Grades			
2	-.7759043	.1467345	0.000
3	-1.885514	.1966385	0.000
Self-Reported Timings			
2			
3	-.9753796	.7482458	0.192
	.0750905	.7102262	0.916
Code Work			
2	.0269145	.2374597	0.910
3	-.207775	.1904752	0.275
4	-.4078453	.2059883	0.048

Successful use of single logit regression analysis resulted following factors with significant effect on students' retention in higher education.

1. Gender ( $p = 0.001$ )
2. Race ( $p = 0.176$ )
3. Transfer percentile ( $p = 0.120$ )
4. Desire to finish percentile ( $p = 0.005$ )
5. Receptivity to academic assistance ( $p = 0.014$ )
6. Receptivity to financial guidance ( $p = 0.032$ )
7. Receptivity to personal counseling ( $p = 0.0152$ )
8. Receptivity to social enrichment ( $p = 0.171$ )
9. Academic stress percentile ( $p = 0.000$ )
10. Attitude towards educators percentile ( $p = 0.019$ )
11. Family emotional support percentile ( $p = 0.001$ )
12. Self-reported College Prep Percentile ( $p = 0.000$ )
13. Math and Science confidence percentile ( $p = 0.089$ )
14. College Prep Percentile ( $p = 0.000$ )
15. High School Grade Percentile ( $p = 0.000$ )
16. Intellectual interest percentile ( $p = 0.075$ )
17. Study habit percentile ( $p = 0.000$ )
18. Desire to transfer percentile ( $p = 0.095$ )
19. Max ACT and SAT score ( $p = 0.000$ )
20. High school GPA ( $p = 0.000$ )
21. Class percentile ( $p = 0.000$ )

- 22. PELL ( $p = 0.000$ )
- 23. Major code ( $p = 0.021$ )
- 24. Father's education ( $p = 0.000$ )
- 25. Mother's education ( $p = 0.000$ )
- 26. Transfer desire ( $p = 0.073$ )
- 27. Senior year grades ( $p = 0.000$ )
- 28. Self-reported timing of decision ( $p = 0.192$ )
- 29. Work ( $p = 0.048$ )

## First logistic regression run:

The first logistic regression model was built using 29 predictor variables demonstrated to be significant factors affecting students' retention in higher education in single logit models (Output window 1).

## OUTPUT WINDOW 1: LOGISTIC REGRESSION OUTPUT WINDOW OF INITIAL REGRESSION MODEL WITH ALL 29 SIGNIFICANT FACTORS

```
. logit allgraduateanddropouts i.CodedGender i.CodedRace TransferPercentile DesiretoFinishPercentile ReceptivitytoAcademicAssistan Race
> pivitytoFinancialGuidanc ReceptivitytoPersonalCounsel ReceptivitytoSocialEnrichment AcademicStresspercentile AttitudeTowardEducatorper
> ce FamilyEmotionalSupportpercen SenseofFinancialSecurityper Selfreportedcollepprepperc MathandScienceConfidenceper collegepreppercenile
> Highschoolgradespercentile IntellectualInterestpercentil StudyHabitspercentile DesiretoTransferpercentile MathACTSATScore HSDPA Classper
> cent i.CodedPELL i.MAJORCODE i.FathersEducation i.MothersEducation i.TransferDesire i.CodedSeniorYeargrades i.CodedSelfreportedtimingofde
> ofde i.CodedWork

Iteration 0: log likelihood = -640.46799
Iteration 1: log likelihood = -533.96314
Iteration 2: log likelihood = -532.99449
Iteration 3: log likelihood = -532.99329
Iteration 4: log likelihood = -532.99329

Logistic regression              Number of obs      =      924
                                LR chi2(44)         =     214.95
                                Prob > chi2          =     0.0000
                                Pseudo R2           =     0.1678

Log likelihood = -532.99329
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1.CodedGender	-.1148592	.1824295	0.63	0.529	-.2426961 .4724145
CodedRace					
2	-.0875484	.2195145	-0.40	0.490	-.5179889 .3426921
3	.2555274	.3411757	0.75	0.454	-.4131446 .9242195
4	-.1544845	.3165971	0.49	0.626	-.4660343 .7750034
TransferPercentile	.0792787	.0453987	1.75	0.081	-.0097011 .1482586
DesiretoFinishPercentile	.0031679	.0043273	0.73	0.464	-.0053134 .0116491
ReceptivitytoAcademicAssistan	.0048485	.003766	1.29	0.198	-.0025327 .0122297
ReceptivitytoFinancialGuidanc	-.0043738	.0037663	-1.16	0.246	-.0117556 .003008
ReceptivitytoPersonalCounsel	.0030757	.0037827	0.81	0.416	-.0043381 .0104896
ReceptivitytoSocialEnrichment	.001913	.0035841	0.53	0.594	-.0051117 .0089978
AcademicStresspercentile	.0103848	.0095361	1.09	0.276	-.0083056 .0290753
AttitudeTowardEducatorperce	.0015039	.0038588	0.39	0.697	-.0060592 .0090669
FamilyEmotionalSupportpercen	.0025972	.0028296	0.92	0.359	-.0029467 .0081431
SenseofFinancialSecurityper	.0031954	.0033298	0.96	0.337	-.0033331 .0097218
Selfreportedcollepprepperc	-.0145182	.0184699	-0.79	0.432	-.0507186 .0216822
MathandScienceConfidenceper	-.0034925	.0041789	-0.84	0.403	-.0116829 .0048979
collegepreppercenile	.0175384	.0183422	0.96	0.339	-.0186107 .0536895
Highschoolgradespercentile	-.003788	.0185497	-0.20	0.838	-.0401448 .0325688
IntellectualInterestpercentil	-.0070231	.0044276	-0.16	0.874	-.0093081 .0079748
StudyHabitspercentile	.0099727	.0041483	2.39	0.022	.0014502 .0184952
DesiretoTransferpercentile	-.0788595	.0461523	-1.71	0.088	-.1693164 .0115974
MathACTSATScore	.0489486	.0293549	2.35	0.019	.011414 .1264832
HSDPA	.7274412	.1717118	2.48	0.007	.1348898 1.259987
Classpercent	-.0095981	.0054415	-1.76	0.078	-.0202433 .0010671
1.CodedPELL	-.1914604	.1752979	-1.09	0.275	-.3505979 .1521171
MAJORCODE					
2	-.4777725	.4974985	-1.36	0.173	-1.450305 .0973667
3	-.6030861	.4274744	-1.41	0.158	-1.44092 .2347483
4	-.3327478	.4191881	-0.79	0.427	-1.154341 .4888457
5	-.6112511	.4207745	-0.98	0.325	-1.827947 .6054447
6	-.2267757	.4326317	-0.52	0.600	-1.074718 .6213668
FathersEducation					
4	-.1788454	.2084599	-0.86	0.391	-.5874393 .2297086
5	-.0880454	.2308717	-0.38	0.703	-.5405456 .3444547
6	-.1747306	.3021288	-0.58	0.563	-.7468923 .4174331
MothersEducation					
4	.3835088	.208437	1.84	0.066	-.0250202 .7920377
5	.7604879	.2328057	3.27	0.001	.304197 1.216779
6	.5616206	.2844398	1.97	0.048	-.0041289 1.119112
1.TransferDesire	-.210608	.3385981	-0.62	0.534	-.874248 .453032
CodedSeniorYearGrades					
2	-.4181799	.4867261	-0.86	0.390	-1.372146 .5357858
3	-1.179684	1.098039	-1.07	0.283	-3.331801 .9724328
CodedSelfreportedtimingofde					
2	-1.028926	.8917141	-1.15	0.249	-2.776653 .7188019
3	-.3060578	.8519778	-0.36	0.719	-1.975904 1.363788
CodedWork					
2	.4741364	.2979535	1.59	0.112	-.1098418 1.058115
3	.2909769	.5315627	1.16	0.247	-.2020768 .7840307
4	.0879215	.2664551	0.33	0.741	-.434321 .610164
_cons	-4.4577717	2.643918	-1.69	0.092	-9.639701 .7242665

Since the p-value is less than 0.05, the model itself is statistically significant. The least significant variable in the model is intellectual Interest with a p-value of 0.974. It is removed and regression ran again.

## Output window 2: Logistic regression output window when intellectual interest is dropped

```
. logit allgraduatesanddropouts i.CodifiedGender i.CodedRace TransferPercentile DesiretoFinishPercentile ReceptivitytoAcademicAssistan Rece
> ptivitytoFinancialGuidanc ReceptivitytoPersonalCounseli ReceptivitytoSocialEnrichment AcademicStresspercentile AttitudeTowardEducatorsper
> ce FamilyEmotionalSupportpercen SenseofFinancialSecuritytyper Selfreportedcollegeprepperperc MathandScienceConfidenceper collegeprepperpercentile
> Highschoolgradespercentile StudyHabitspercentile DesiretoTransferpercentile MaxACTSATscore HSGPA Classpercent i.CodedPELL i.MAJORCODE i
> .FathersEducation i.MothersEducation i.TransferDesire i.CodedSeniorYearGrades i.CodedSelfreportedtimingofde i.CodedWork
```

```
Iteration 0: log likelihood = -640.46799
Iteration 1: log likelihood = -533.97378
Iteration 2: log likelihood = -533.00911
Iteration 3: log likelihood = -533.0059
Iteration 4: log likelihood = -533.0059
```

```
Logistic regression      Number of obs      =      924
                        LR chi2(43)      =     214.92
                        Prob > chi2      =     0.0000
Log likelihood = -533.0059      Pseudo R2      =     0.1678
```

allgraduatesanddropouts	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.CodifiedGender	.1106698	.1805129	0.61	0.540	-.243129	.4644686
CodedRace						
2	-.0879303	.2195254	-0.40	0.689	-.5181922	.3423315
3	.2588835	.3405112	0.76	0.447	-.4085062	.9262733
4	.1549765	.3165907	0.49	0.624	-.46553	.7754829
TransferPercentile	.0805359	.0447918	1.80	0.072	-.0072545	.1683262
DesiretoFinishPercentile	.003477	.0038652	0.90	0.368	-.0040987	.0110527
ReceptivitytoAcademicAssistan	.0048066	.0037571	1.28	0.201	-.0025571	.0121703
ReceptivitytoFinancialGuidanc	-.0043796	.0037662	-1.16	0.245	-.0117613	.0030021
ReceptivitytoPersonalCounseli	.0030235	.003768	0.80	0.422	-.0043617	.0104086
ReceptivitytoSocialEnrichment	.0018883	.0035807	0.53	0.598	-.0051297	.0089064
AcademicStresspercentile	.0115201	.0063153	1.82	0.068	-.0008577	.0238979
AttitudeTowardEducatorsperce	.0017895	.0034142	0.52	0.600	-.0049022	.0084813
FamilyEmotionalSupportpercen	.0026086	.0028284	0.92	0.356	-.0029349	.0081522
SenseofFinancialSecuritytyper	.0031953	.0033305	0.96	0.337	-.0033324	.009723
Selfreportedcollegeprepperperc	-.0141396	.0183054	-0.77	0.440	-.0500176	.0217383
MathandScienceConfidenceper	-.0032645	.003924	-0.83	0.405	-.0109554	.0044263
collegeprepperpercentile	.0172378	.018233	0.95	0.344	-.0184984	.0529739
Highschoolgradespercentile	-.003735	.0185452	-0.20	0.840	-.0400829	.0326128
StudyHabitspercentile	.0102298	.0040368	2.53	0.011	.0023178	.0181418
DesiretoTransferpercentile	-.0801248	.0455464	-1.76	0.079	-.1693941	.0091444
MaxACTSATscore	.068122	.0288894	2.36	0.018	.0114999	.1247442
HSGPA	.7274165	.2716321	2.68	0.007	.1950273	1.259806
Classpercent	-.0096166	.0054382	-1.77	0.077	-.0202754	.0010421
1.CodedPELL	-.1931475	.1749592	-1.10	0.270	-.5360612	.1497662
MAJORCODE						
2	-.6765721	.4974654	-1.36	0.174	-1.651586	.2984421
3	-.5994335	.4268435	-1.40	0.160	-1.436031	.2371644
4	-.3276534	.4179475	-0.78	0.433	-1.146815	.4915086
5	-.6069275	.6204511	-0.98	0.328	-1.822989	.6091344
6	-.2245564	.4324112	-0.52	0.604	-1.072067	.622954
FathersEducation						
4	-.1812226	.2079358	-0.87	0.383	-.5887691	.226324
5	-.0901189	.230495	-0.39	0.696	-.5418807	.361643
6	-.1759588	.3020492	-0.58	0.560	-.7679643	.4160467
MothersEducation						
4	.3836786	.2084087	1.84	0.066	-.0247949	.7921522
5	.7601587	.2327766	3.27	0.001	.3039249	1.216392
6	.5617307	.2844388	1.97	0.048	.0042408	1.119221
1.TransferDesire	-.2097114	.3385316	-0.62	0.536	-.8732212	.4537984
CodedSeniorYearGrades						
2	-.4175143	.4866771	-0.86	0.391	-1.371384	.5363553
3	-1.177952	1.097912	-1.07	0.283	-3.32982	.9739167
CodedSelfreportedtimingofde						
2	-1.021847	.8899677	-1.15	0.251	-2.766152	.7224577
3	-.3008429	.8506984	-0.35	0.724	-1.968181	1.366495
CodedWork						
2	.4735874	.2979231	1.59	0.112	-.1103312	1.057506
3	.2894167	.2513521	1.15	0.250	-.2032245	.7820578
4	.0861149	.266235	0.32	0.746	-.4356962	.607926
_cons	-4.590785	2.506873	-1.83	0.067	-9.504166	.3225966

Based on p values the least significant variable is high school grades ( $p = 0.84$ ), and it is dropped in the next run.

**TABLE 2: LIST OF FACTORS AND THEIR P-VALUES THAT HAVE BEEN DROPPED FROM REGRESSION ANALYSIS.**

<b>DROPPED FACTORS</b>	<b>P VALUE</b>	<b>LOG LIKELIHOOD</b>
Intellectual interest percentile	0.974	-533.0058
High school grades	0.84	-533.02612
Social enrichment	0.596	-533.16683
Attitude toward educators	0.580	-533.32007
Gender	0.478	-533.57122
Transfer desire	0.545	-532.24222
Race	0.835	-540.24122
Fathers education	0.820	-542.57883
Family emotional support	0.445	-542.35999
Receptivity to personal counseling	0.456	-543.14832
Desire to finish percentile	0.440	-543.44647
Self-reported college prep percentile	0.412	-543.80637
College prep percentile	0.449	-544.09191
Sense of financial security percentile	0.293	-544.64501
Academic stress percentile	0.307	-545.16578
Work	0.919	-546.02649
Self-reported timing of	0.528	-549.79169

decision		
Major	0.635	-552.71967
Receptivity to financial guidance	0.635	-553.97281
Receptivity to academic assistance	0.146	-555.03672
Transfer percentile	0.139	-556.52864
Desire to transfer percentile	0.118	-557.75208
Max act sat scores	0.078	-562.57323

## Final model with all significant factors:

```
. logit allgraduatesanddropouts MathandScienceConfidenceper StudyHabitspercentile HSGPA Classpercent i.CodedPELL i.MothersEducation i.Code
> dSeniorYearGrades
```

```
Iteration 0: log likelihood = -650.86467
Iteration 1: log likelihood = -563.02088
Iteration 2: log likelihood = -562.57396
Iteration 3: log likelihood = -562.57323
Iteration 4: log likelihood = -562.57323
```

```
Logistic regression              Number of obs   =       939
                                LR chi2(10)      =      176.58
                                Prob > chi2       =       0.0000
Log likelihood = -562.57323      Pseudo R2     =       0.1357
```

allgraduatesanddropouts	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
MathandScienceConfidenceper	-.0072021	.0029428	-2.45	0.014	-.0129698	-.0014343
StudyHabitspercentile	.0061409	.0025193	2.44	0.015	.0012031	.0110786
HSGPA	.7831528	.2560111	3.06	0.002	.2813802	1.284925
Classpercent	-.0144057	.0048178	-2.99	0.003	-.0238483	-.0049631
1.CodedPELL	-.3251892	.1517781	-2.14	0.032	-.6226688	-.0277096
MothersEducation						
4	.3764006	.1832312	2.05	0.040	.017274	.7355271
5	.7861789	.1990692	3.95	0.000	.3960105	1.176347
6	.6356455	.2437468	2.61	0.009	.1579106	1.11338
CodedSeniorYearGrades						
2	-.3235641	.1755024	-1.84	0.065	-.6675424	.0204142
3	-.9321909	.2614496	-3.57	0.000	-1.444623	-.4197592
_cons	-2.080515	1.044994	-1.99	0.046	-4.128665	-.0323656



**TABLE 3: FINAL LIST OF FACTORS INFLUENCING RATE OF GRADUATION**

<b>VARIABLE</b>	<b>P VALUE</b>
Math and science confidence percentile	0.014
Study habits percentile	0.015
HSGPA	0.002
Class percent	0.003
PELL	0.032
Mother's education	
4	0.040
5	0.000
6	0.009
Senior year grades	
2	0.065
3	0.000

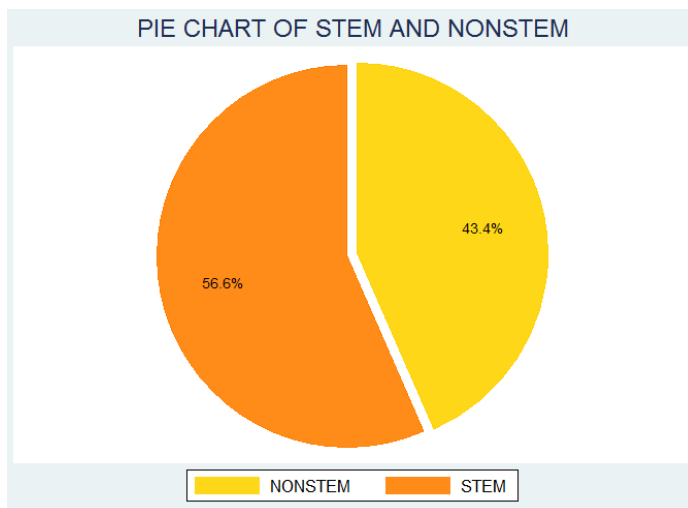
## Summary

- 50.4 % of the students graduated with either a 4-year degree or a community college degree, while 49.6% dropped out.
- Student success or failure in higher education is affected by the following predictors:
  - Math and science confidence percentile
  - study habits percentile
  - HSGPA
  - Class percent
  - PELL
  - Mothers' education
  - Senior year grades.

## CHAPTER IV: FACTORS AFFECTING RATE GRADUATION IN STEM MAJORS

The second objective of this study was to identify factors that impact retention to graduation with a STEM degree. This chapter will present the exploration and analysis related to answering the second research question. What factors impact retention to graduation in STEM degree program? We will first present the exploration of the outcome variable, Graduation with a STEM degree, indicating the student graduated with STEM degree in some manner, versus all predictor variables. Then the formal development of a logistic regression model for Graduation with STEM degree, based on these predictors will be developed.

Here I divided graduates in two groups 1) Graduated with STEM degree 2) Graduated with NONSTEM degree.



**FIGURE 32: DISTRIBUTION OF GRADUATES BASED ON MAJOR**

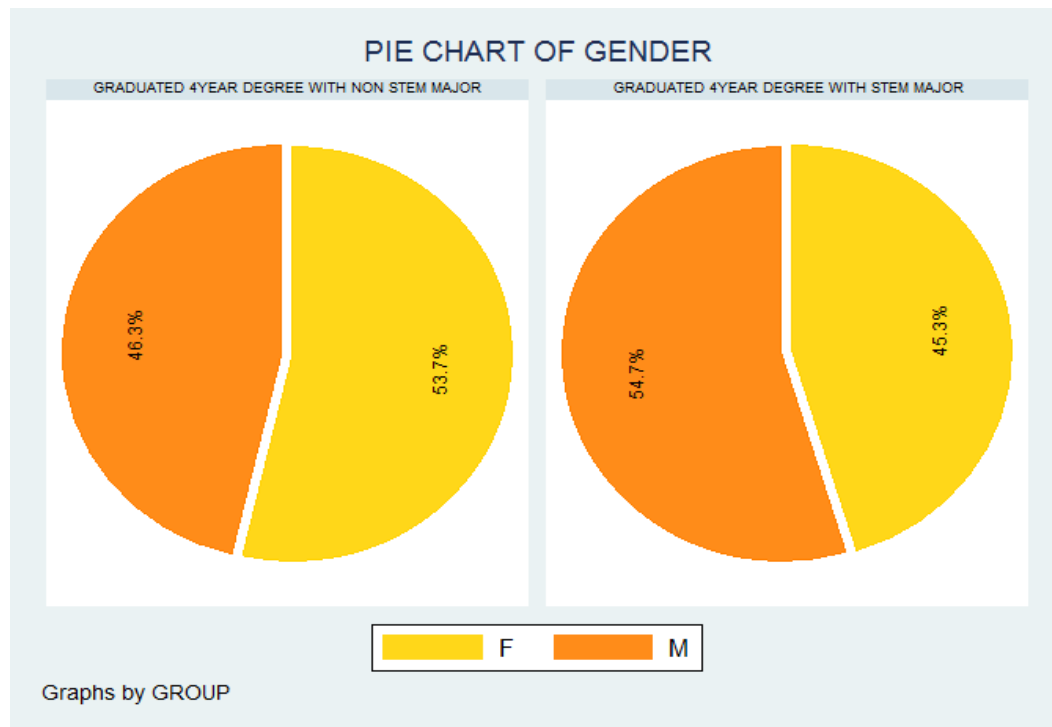
**Inference:** There are more STEM graduates than NONSTEM graduates.

## Exploration

Some exploratory graphs were prepared before building the statistical model to guide the process and determine how to proceed. The exploration begins with identification of important factors affecting student success in STEM degree.

The following pie charts and Box and whisker plots address the two primary groups: 1) Students that graduated with STEM degree and 2) Students that graduated with NONSTEM degree. For each of the variables the pie charts then indicate the proportions of students within the primary categories that satisfy the variables, options for example, the pie charts for the senior year grades variable, Figure 36 illustrates the proportion of students in 1) and 2) of senior year grades, about 15% more STEM students graduated with A average in Senior year.

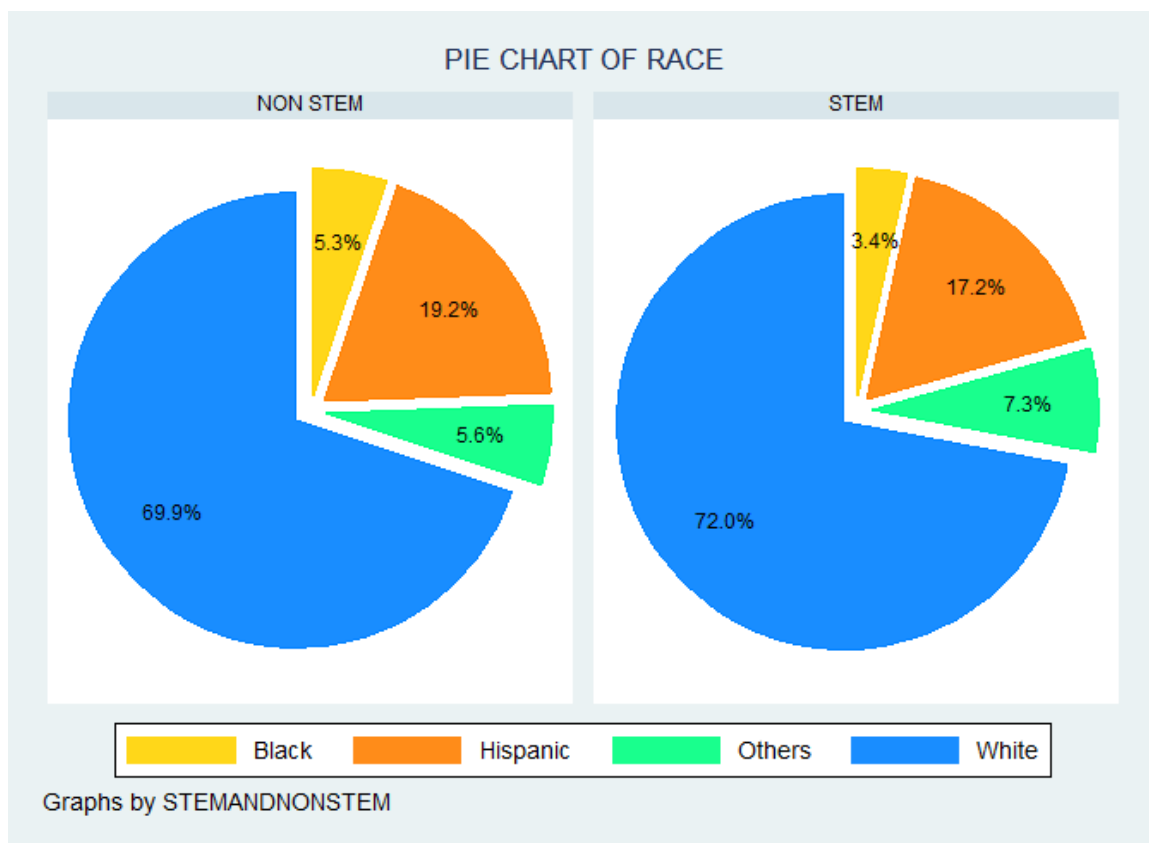
## GRAPHS FOR STEM AND NONSTEM



**FIGURE 33: PIE CHART OF GENDER**

1) Among 4-year degree graduates with non-STEM major 53.7% are female and 46.3% are male. 2) Among 4-year degree graduates with STEM major 45.3% are female and 54.7% male.

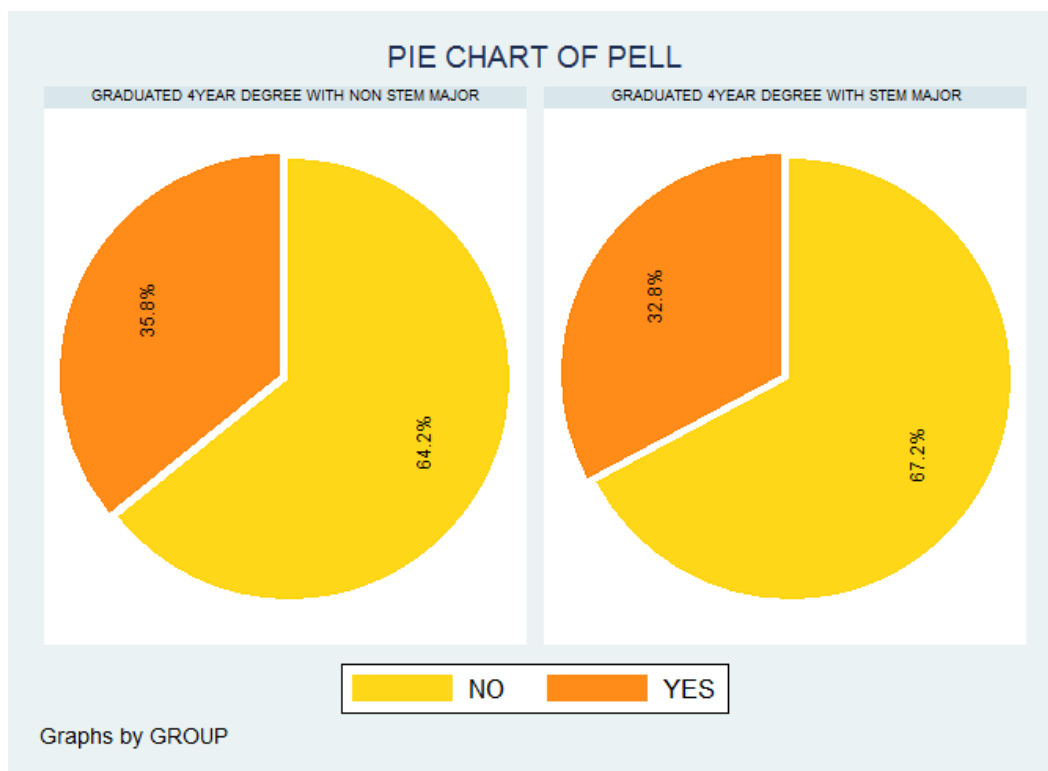
**Inference:** Greater percentage of males (54.4%) graduated with STEM major compared to Females (46.3%).



**FIGURE 34: PIE CHART OF RACE**

- 1) 4-year degree graduates with non-STEM major: 70.6% are white, 1.8% are Asian, 6.0% are Black, 17.9% are Hispanic, 0.6% are Native Amer, 2.3% are others.
- 2) among 4-year degree graduates with STEM major: 72.0% are White, 1.7% are Asian, 3.4% are Black, 17.2% are Hispanic, 2.2% are Native Amer, 3.4% are other.

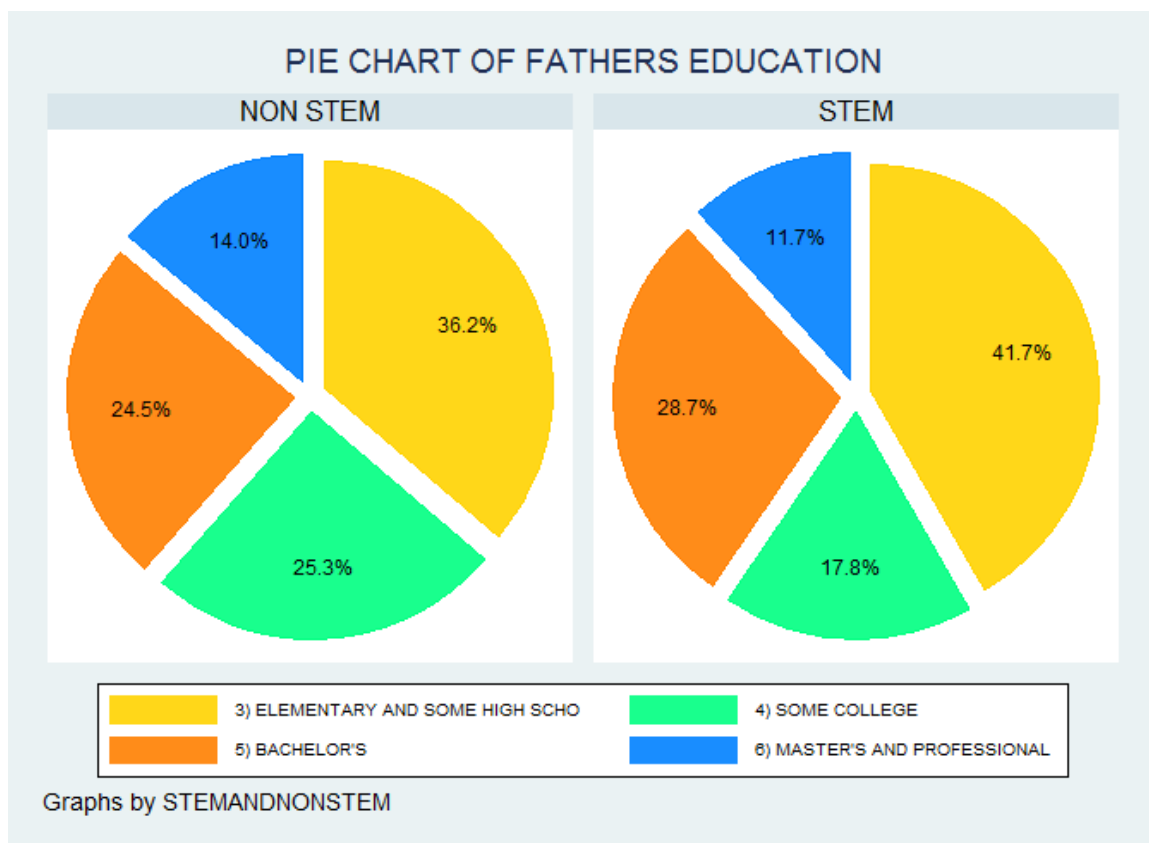
**Inference:** Similar racial distribution is observed among STEM and non-STEM graduates.



**FIGURE 35: PIE CHART OF PELL**

1) Among 4-year graduates with non-STEM major: 35.8% received PELL, and 64.2% not received. 2) Among 4-year graduates with STEM major: 32.8% received PELL, and 67.2% not received PELL.

**Inference:** Similar proportion of STEM and non-STEM graduates received PELL.

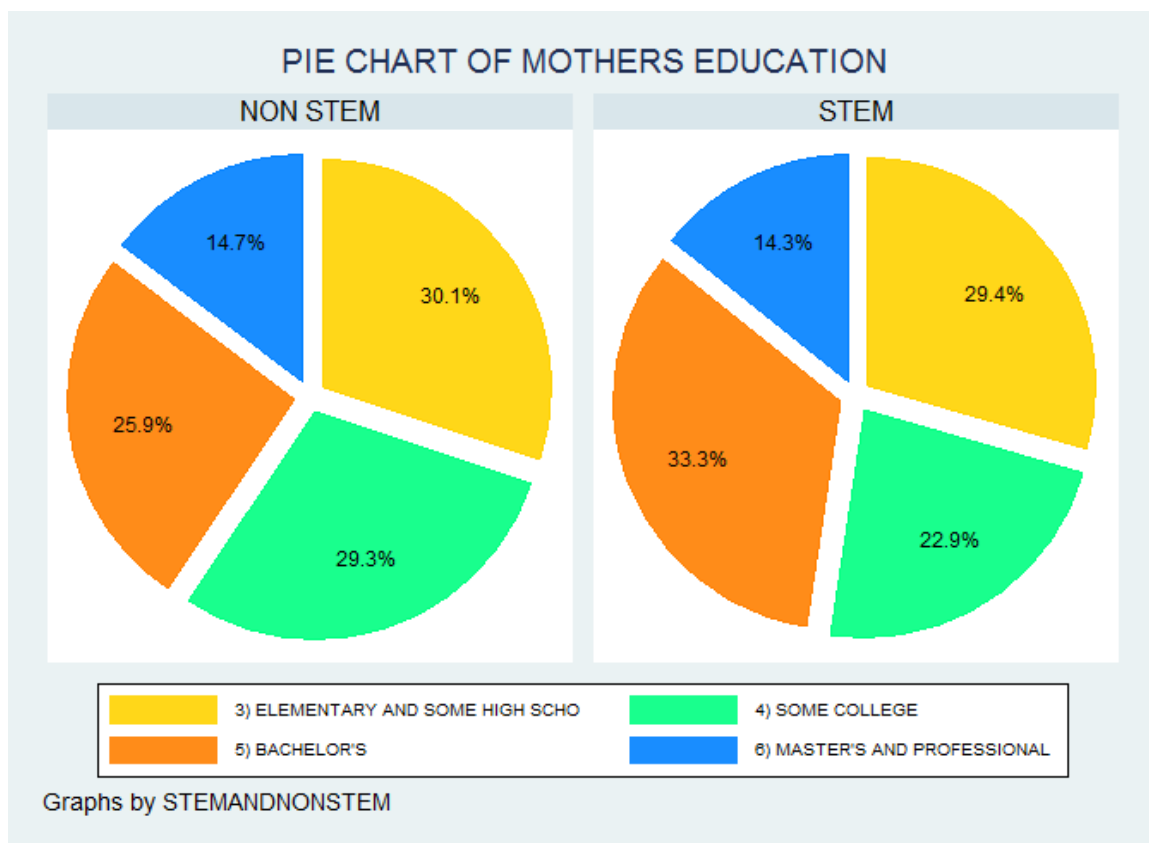


**FIGURE 36: PIE CHART OF FATHERS EDUCATION**

1) Among 4-year graduates with non-STEM major: Fathers education of 26.7% bachelor's, 7.4% elementary, 20.7% high school diploma, 8.8% masters, 6.0% professionals, 24.9% some college, 5.5% some high school. 2) Among 4-year graduates with STEM degree: Fathers education of 28.7% bachelor's, 6.5% elementary, 31.7% high school diploma, 8.5% masters, 5.2% professionals, 17.8% some college, 14.6% some high school.

**Inference:** There is a small difference between the level of parent's education of STEM and non-STEM graduates.

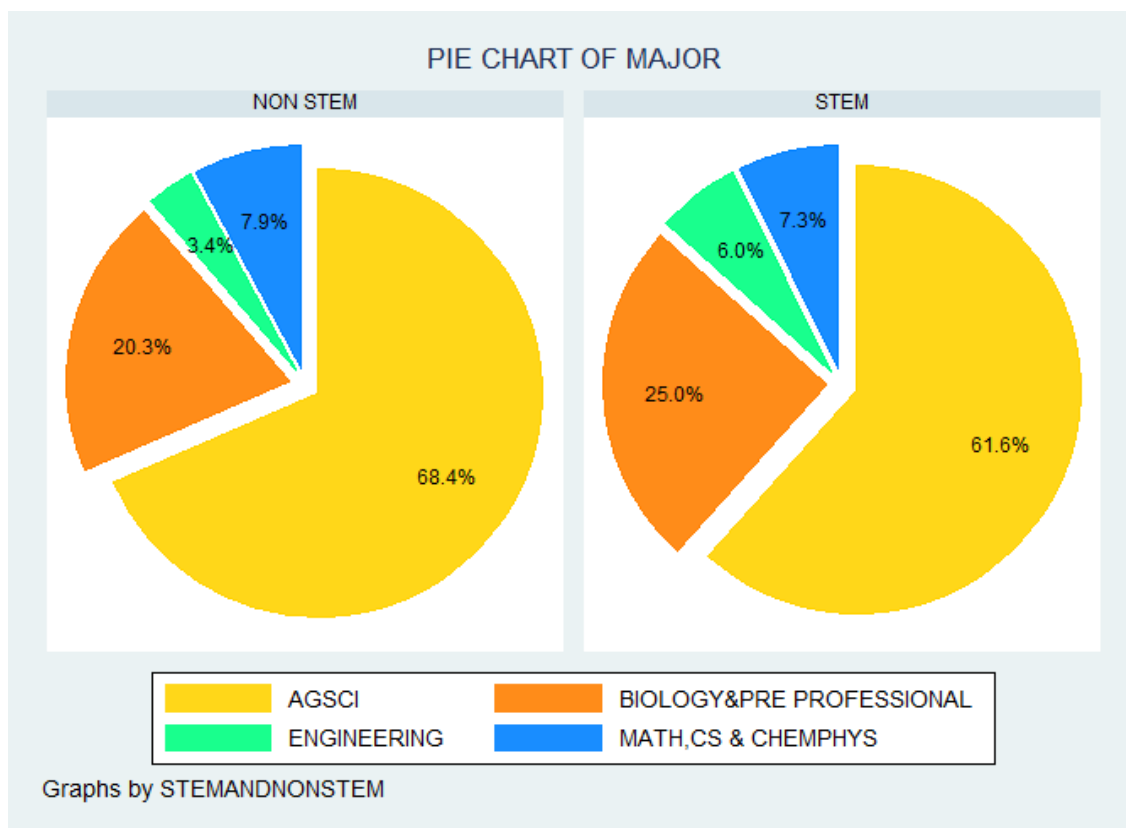




**FIGURE 37: PIE CHART OF MOTHERS EDUCATION**

1) Among 4-year graduates with non-STEM major: Mother's education of 27.5% bachelor's, 5.5% elementary, 17.4% high school diploma, 11.9% masters, 4.1% professionals, 28.0% some college, 5.5% some high school. 2) Among 4-year graduates with STEM degree: Mother's education of 33.3% bachelor's, 6.9% elementary, 19.9% high school diploma, 10.4% masters, 3.9% professionals, 22.9% some college, 2.6% some high school.

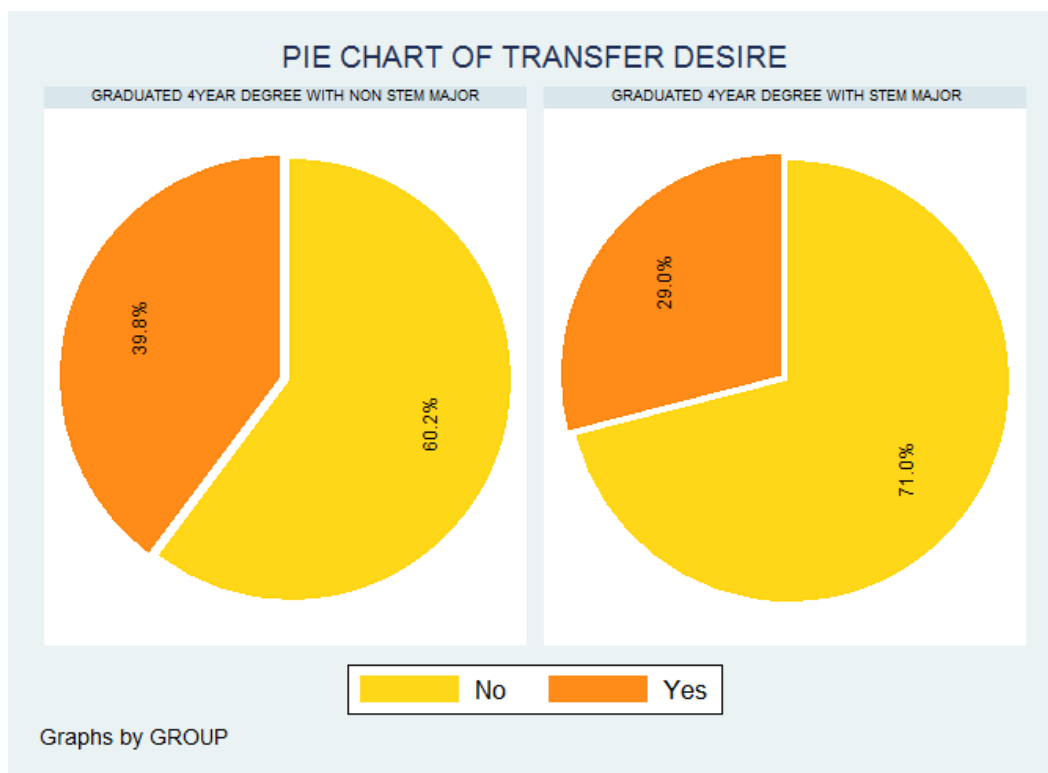
**Inference:** Similar level of mother's education observed between STEM and non-STEM graduates.



**FIGURE 38:PIE CHART OF MAJOR**

1) 4-year degree graduates with non-stem major: 17.9% of agsci, 50.5% of Biology & pre professional, 2.8% of chemphys, 3.7% of cs, 19.7% of engineering and 5.5% of math chosen by students. 2) 4-year degree graduates with stem major: 27.6% of agsci, 34.1% of Biology & pre professional, 3.0% of chemphys, 6.0% of cs, 25.0% of engineering and 4.3% of math chosen by students.

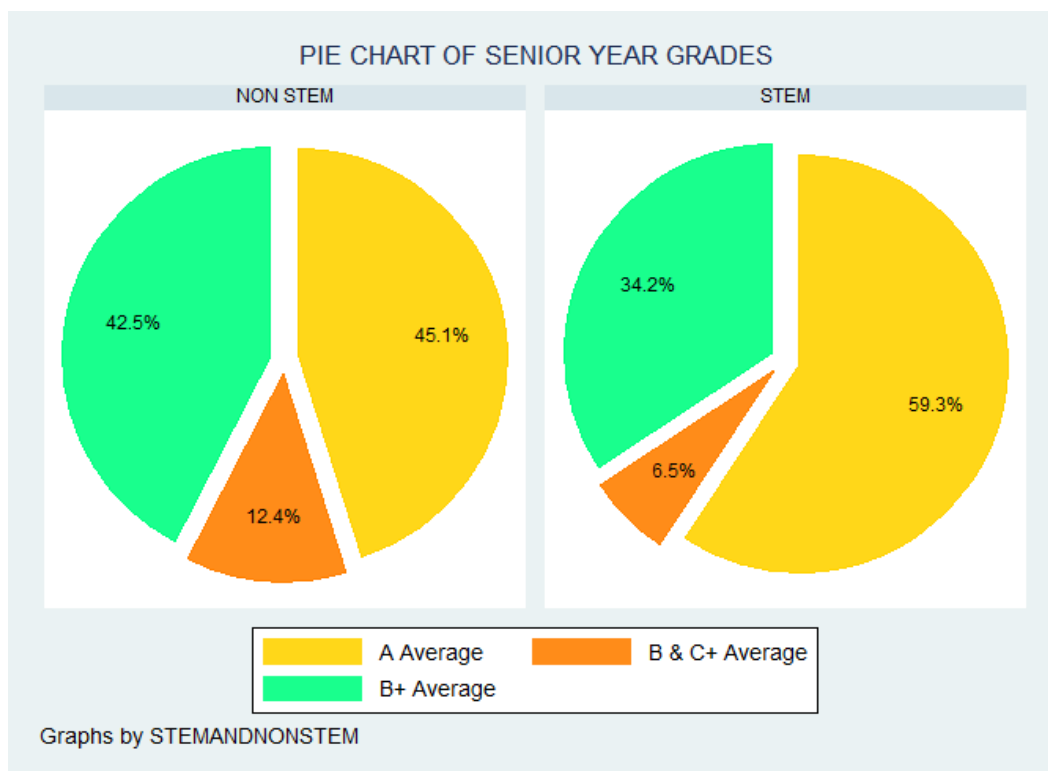
**Inference:** Almost about same percent of STEM and NON STEM graduates chose their majors.



**FIGURE 39: PIE CHART OF TRANSFER DESIRE**

- 1) Among 4-year degree graduates with non-STEM major: 39.8% of students have transfer desire and 60.2% of students don't have transfer desire.
- 2) Among 4-year degree graduates with STEM major: 29.0% of students have transfer desire and 71.0% of students don't have transfer desire.

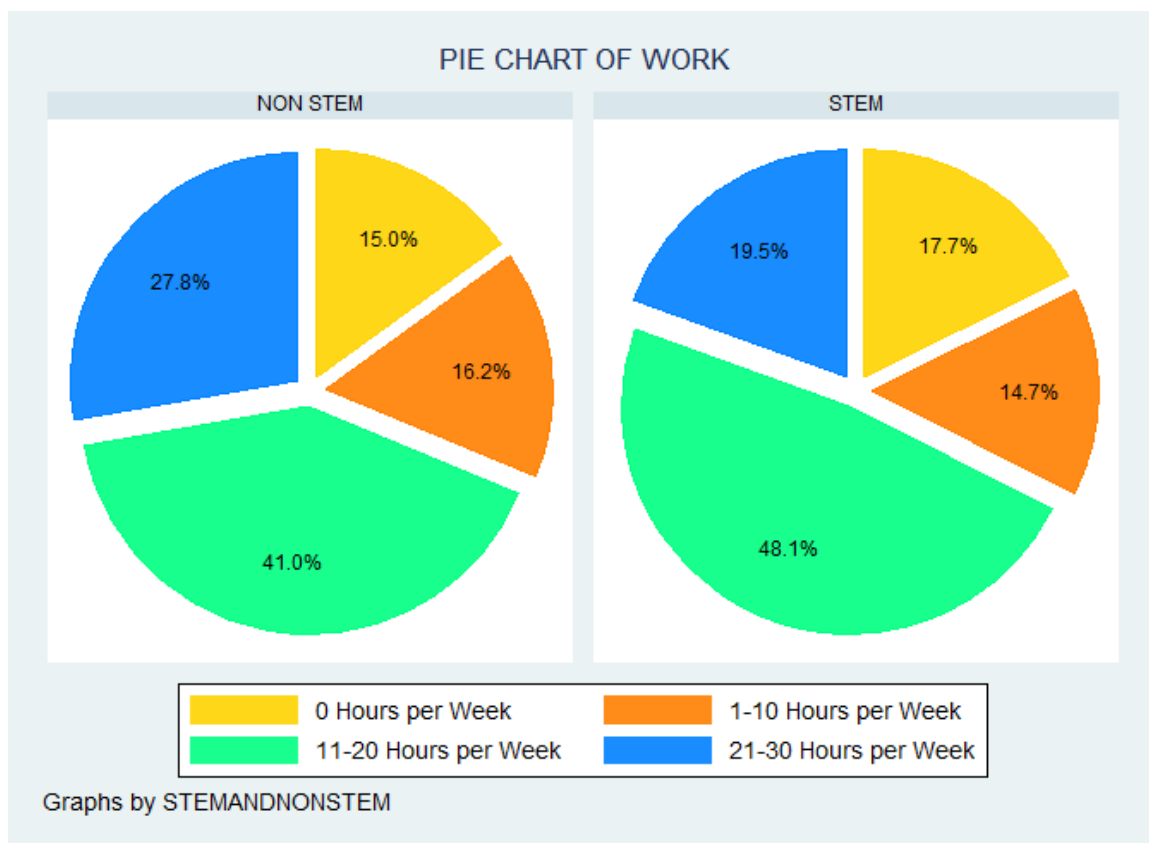
**Inference:** Greater percentage (39.8%) non-STEM graduates expressed transfer desire than STEM graduates (29%).



**FIGURE 40: PIE CHART OF SENIOR YEAR GRADES**

- 1) 4-year degree graduates with non-STEM major: 45.1% had senior year grade A, 42.5% had grade B<sup>+</sup> average, and 12.4% had grade C<sup>+</sup> average, and.
- 2) 4-year degree graduates with STEM major: 59.3% had senior year grade A, 34.2% had grade B<sup>+</sup> average and 6.5% of grades C<sup>+</sup> average.

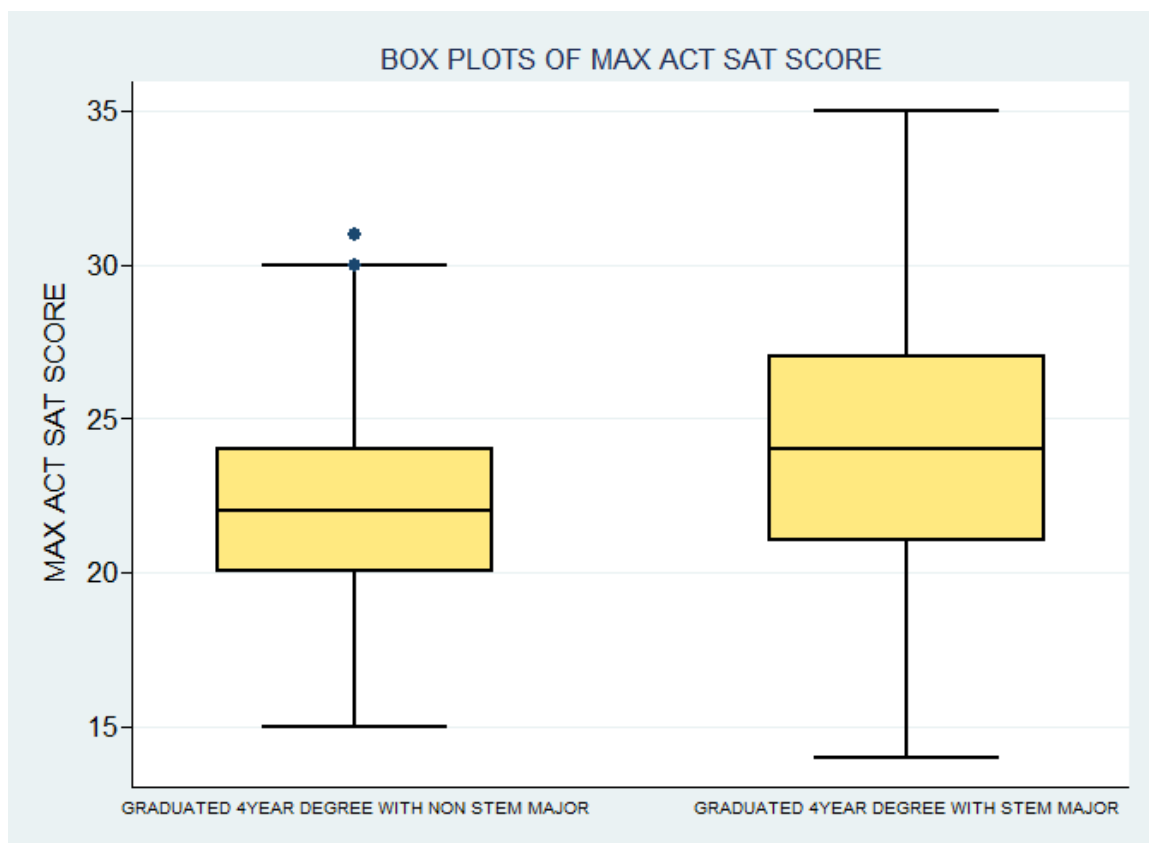
**Inference:** Greater percentage (59.3%) STEM graduates had grade A average compared to non-STEM graduates (45.1%).



**FIGURE 41: PIE CHART OF WORK**

1) Among 4-year graduates with non-STEM degree: 15 % had 0 hours per week, 16.2% had 1-10 hours per week, 41% had 11-20 hours per week, and 27.8% had 21-30 hours per week . 2) Among 4-year graduates with STEM degree: 17.7% had 0 hours per week, 14.7% had 1-10 hours per week, 48.1% had 11-20 hours per week, and 19.5% had 21-30 hours per week.

**Inference:** Less proportion of STEM graduates(19.5%) worked for 21-30 hours per week compared to non-STEM graduates (27.8%).

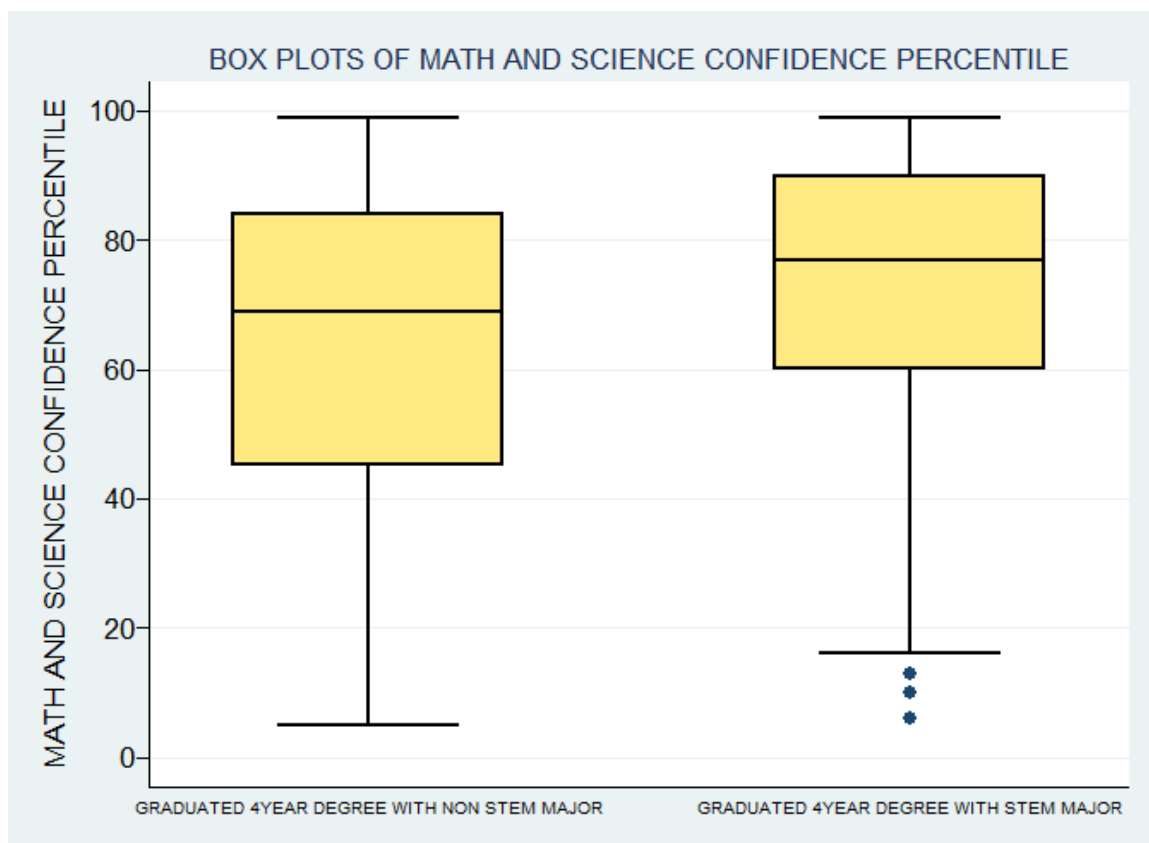


**FIGURE 42: BOX PLOT OF MAX ACT & SAT SCORE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median max ACT & SAT scores of non-STEM graduates is 22 (24-20), and of STEM graduates is 24 (27-21).

**Inference:** The median max ACT & SAT scores of non-STEM graduates is 2 percentile less than that of STEM graduates.

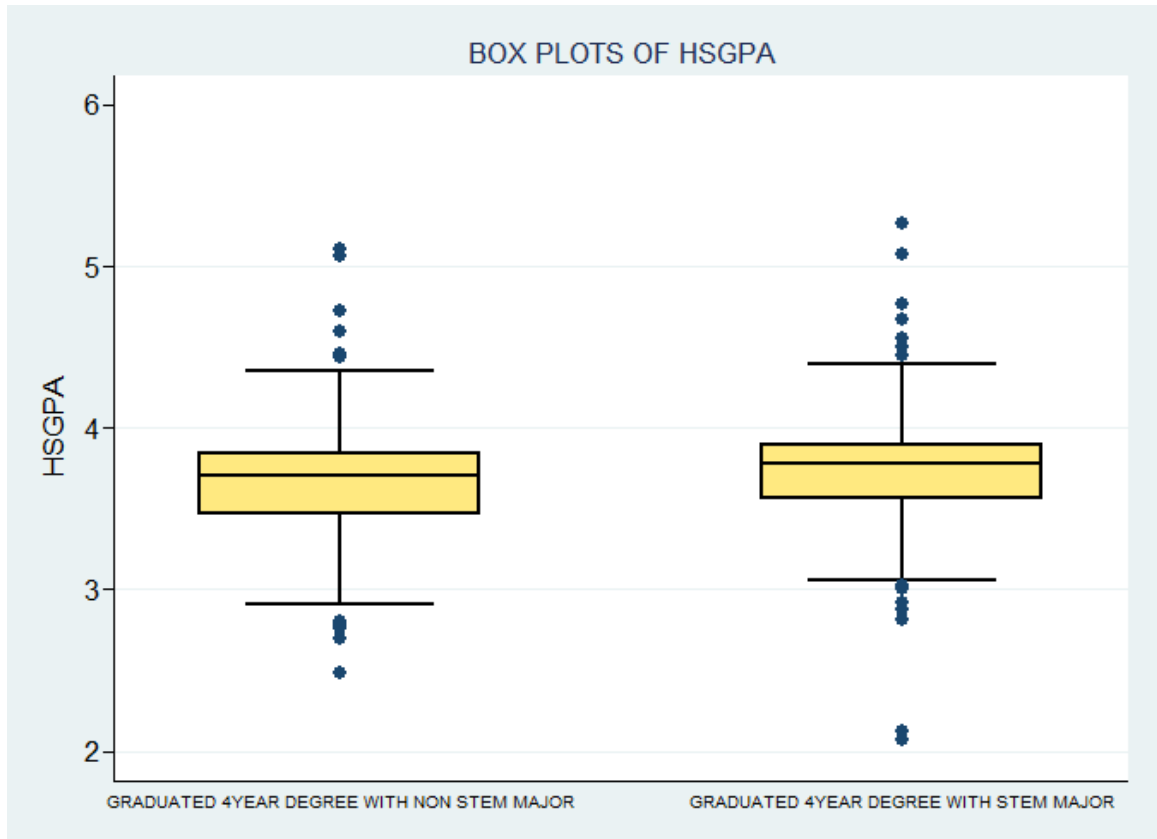


**FIGURE 43: BOX PLOT OF MATH AND SCIENCE CONFIDENCE PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median math & science confidence percentile of non-STEM graduates is 69 (84-45), and of STEM graduates is 77 (90-60).

**Inference:** The median math & science confidence percentile of non-STEM graduates is 8 percentile less than that of STEM graduates.



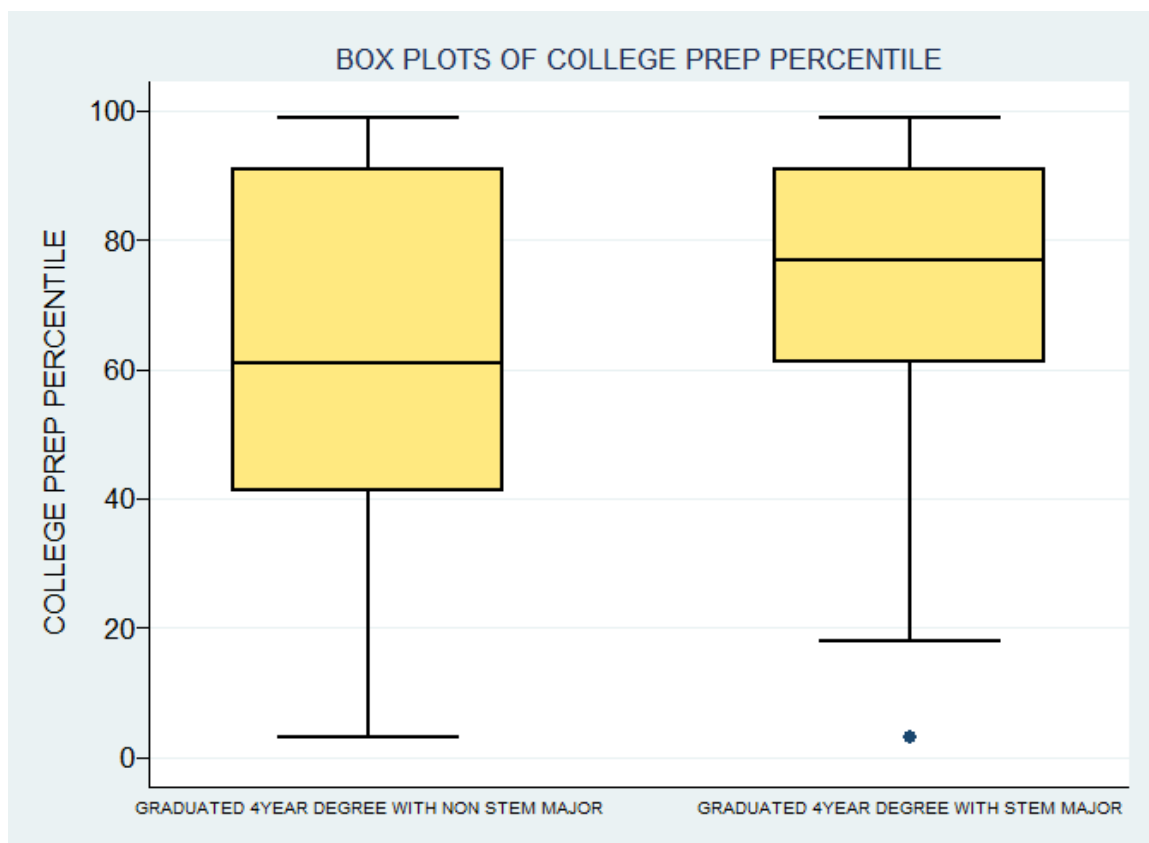
**FIGURE 44: BOX PLOT OF HIGH SCHOOL GPA**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median high school GPA of non-STEM graduates is 3.7 (3.8 - 3.4), and of STEM graduates is 3.8 (3.9 - 3.6).

**Inference:** Median high school GPA of STEM graduates is 0.1 percentile greater than that of non-STEM graduates.



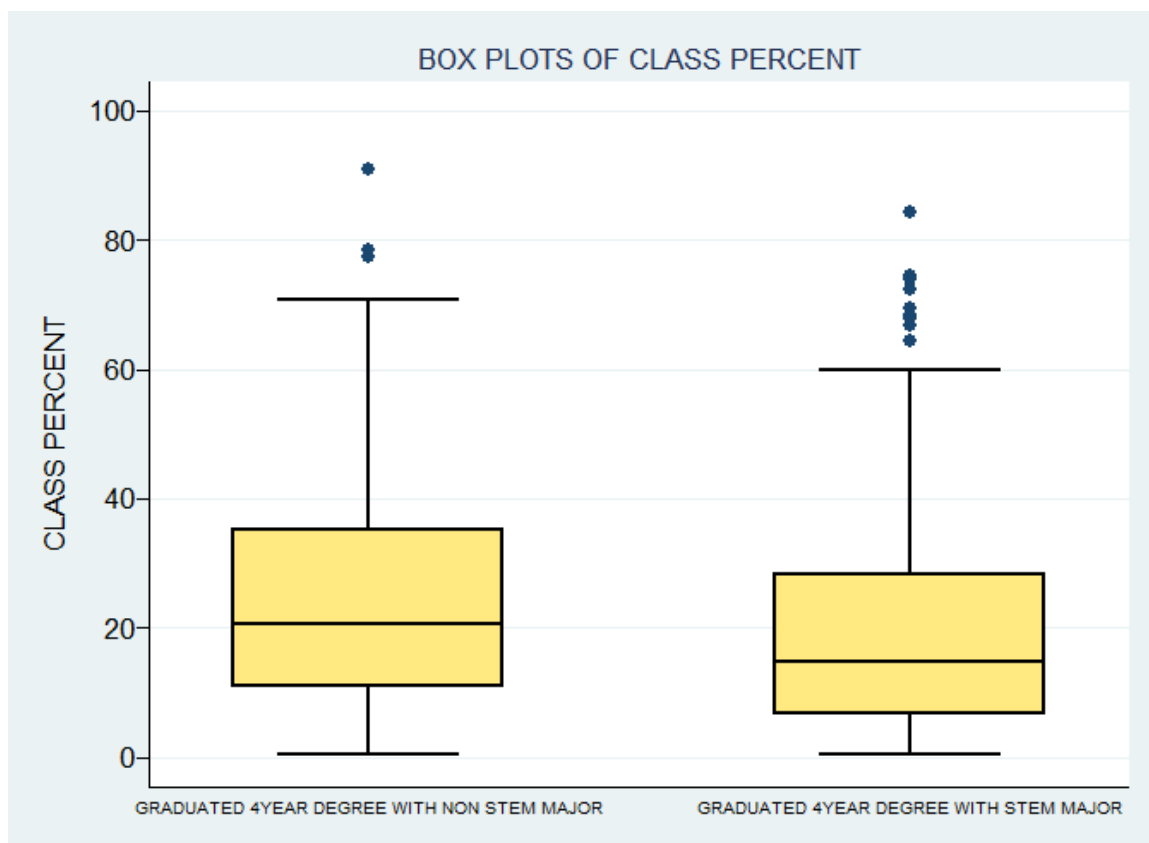


**FIGURE 45: BOX PLOTS OF COLLEGE PREP PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median college prep percentile of non-STEM graduates is 61(77-41), and of STEM graduates is 77 (91-61).

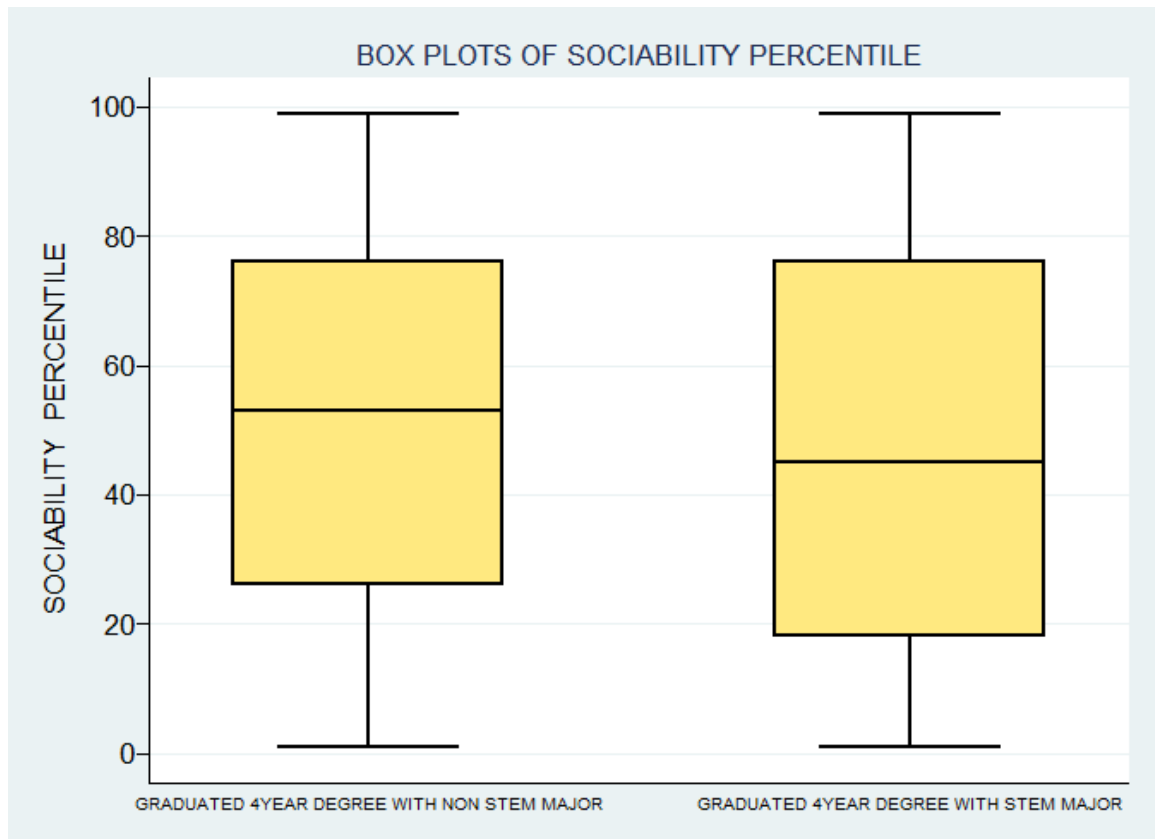
**Inference:** The median college prep percentile of STEM graduates is 16 percentile greater than the non-STEM students.



**FIGURE 46: BOX PLOTS OF CLASS PERCENT**

The median class percent of non-STEM graduates is 22, and of STEM graduates is 15.

**Inference:** The median class percent of non-STEM graduates is 7 percentile less than the STEM graduates.

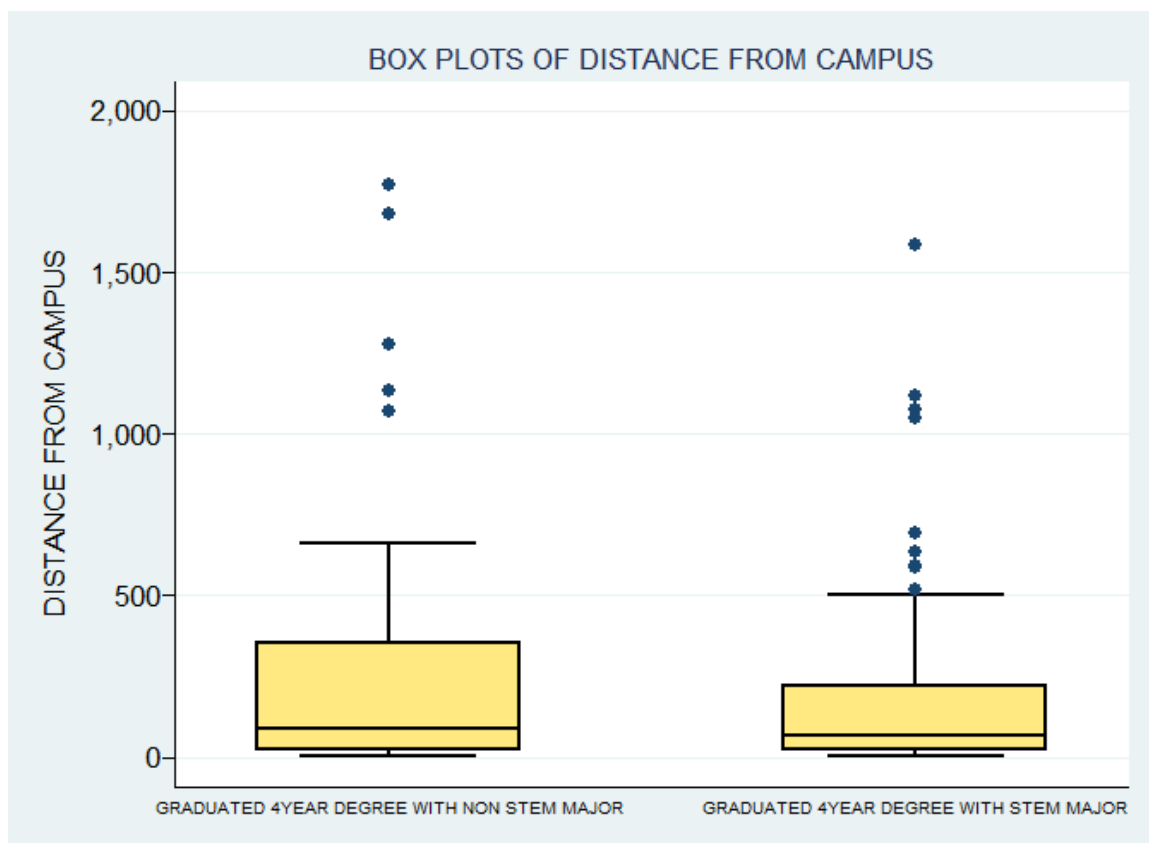


**FIGURE 47: BOX PLOTS OF SOCIABILITY PERCENTILE**

Data are median with the third quartile (Q3) and first quartile (Q1).

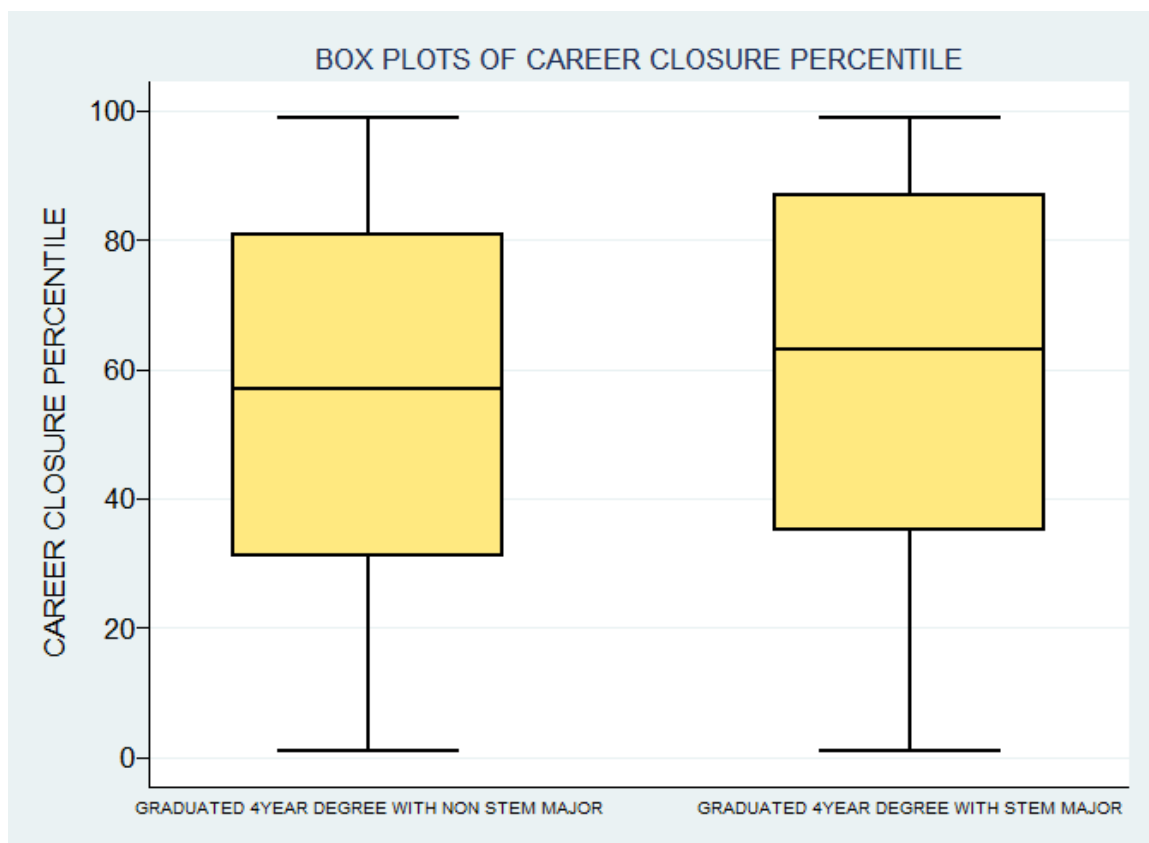
The median sociability percentile of non-STEM graduates is 55 (76-27), and of STEM graduates is 45 (76-18).

**Inference:** The median sociability percentile of non-STEM graduates is 10 percentile less than the STEM students.



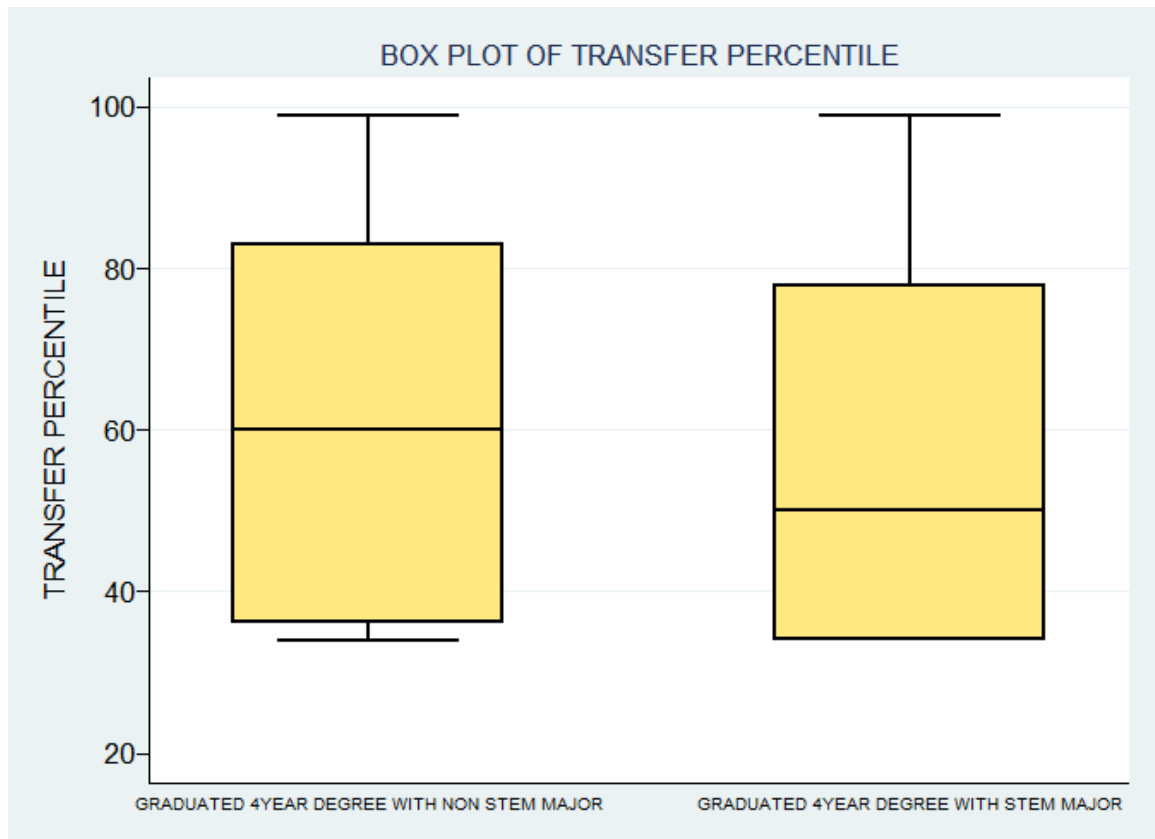
**FIGURE 48: BOX PLOTS OF DISTANCE FROM CAMPUS**

**Inference:** The median of Distance from campus for STEM major graduates is 83 and median of Distance from campus for STEM major graduates is 67. The median STEM graduates is 16 Percentile higher than NONSTEM students.



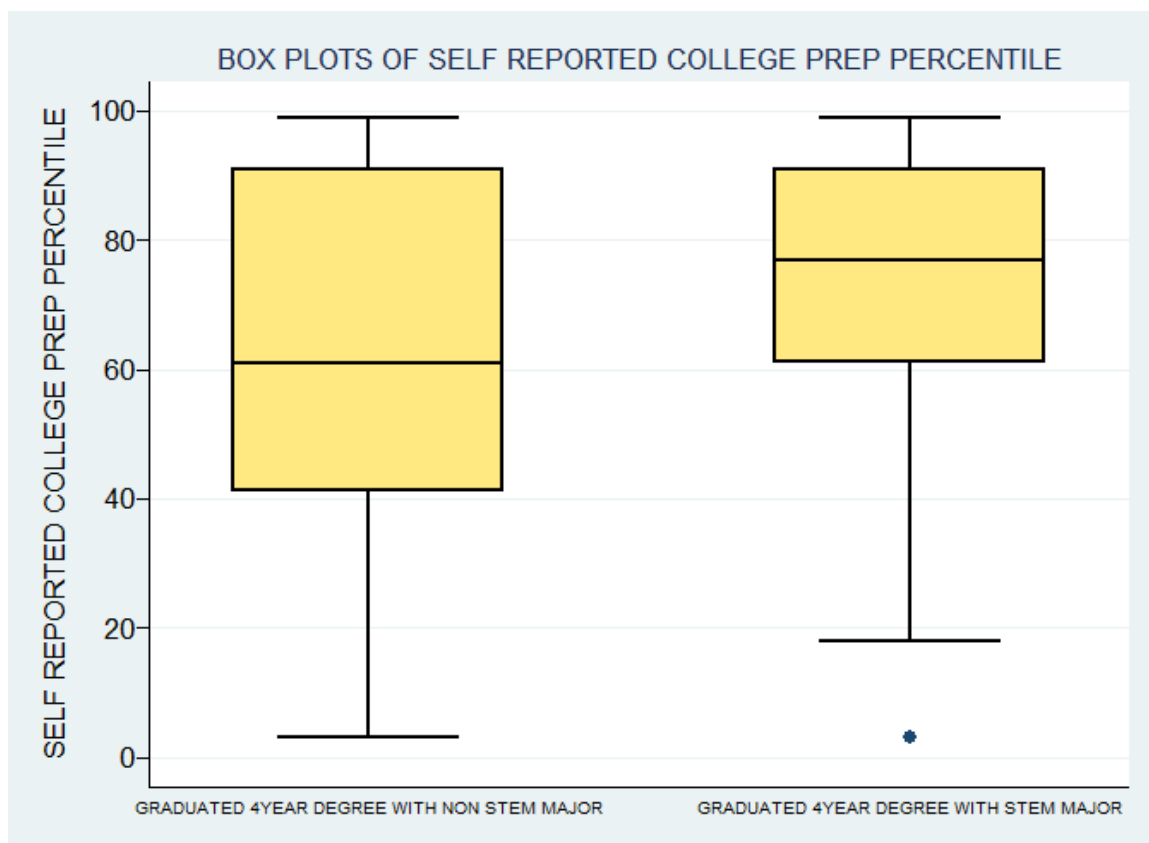
**FIGURE 49: BOX PLOTS OF CAREER CLOSURE PERCENTILE**

**Inference:** The median of Career closure percentile for STEM major graduates is 53 and median of Career closure percentile for STEM major graduates is 83. The median STEM graduates is 30 Percentile higher than NONSTEM students.



**FIGURE 50: BOX PLOT OF TRANSFER PERCENTILE**

**Inference:** The median of Transfer percentile for NONSTEM major graduates is 60 and median of Transfer percentile for STEM major graduates is 50. The median NONSTEM graduates is 10 Percentile higher than STEM students.

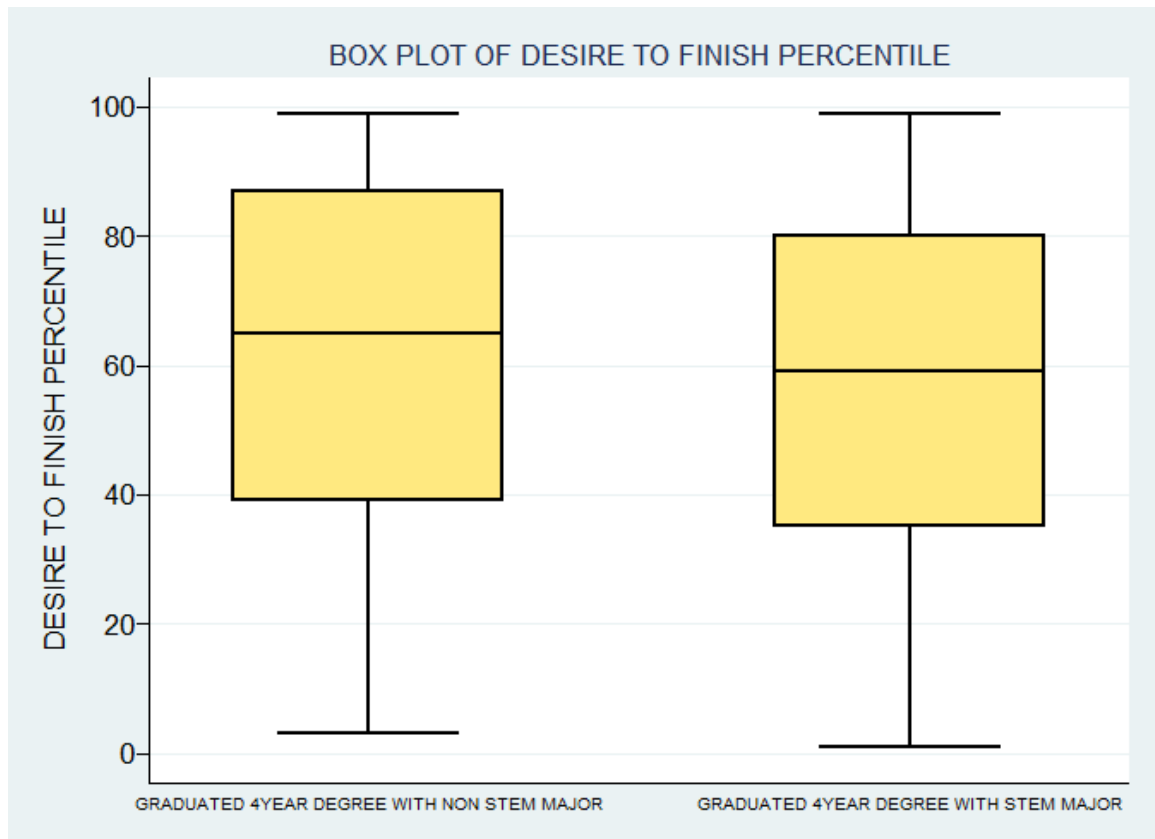


**FIGURE 51: BOX PLOTS OF SELF REPORTED COLLEGE PREP**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median self-reported college prep for non-STEM graduates is 61(77-41), and for STEM graduates is 71(91-61).

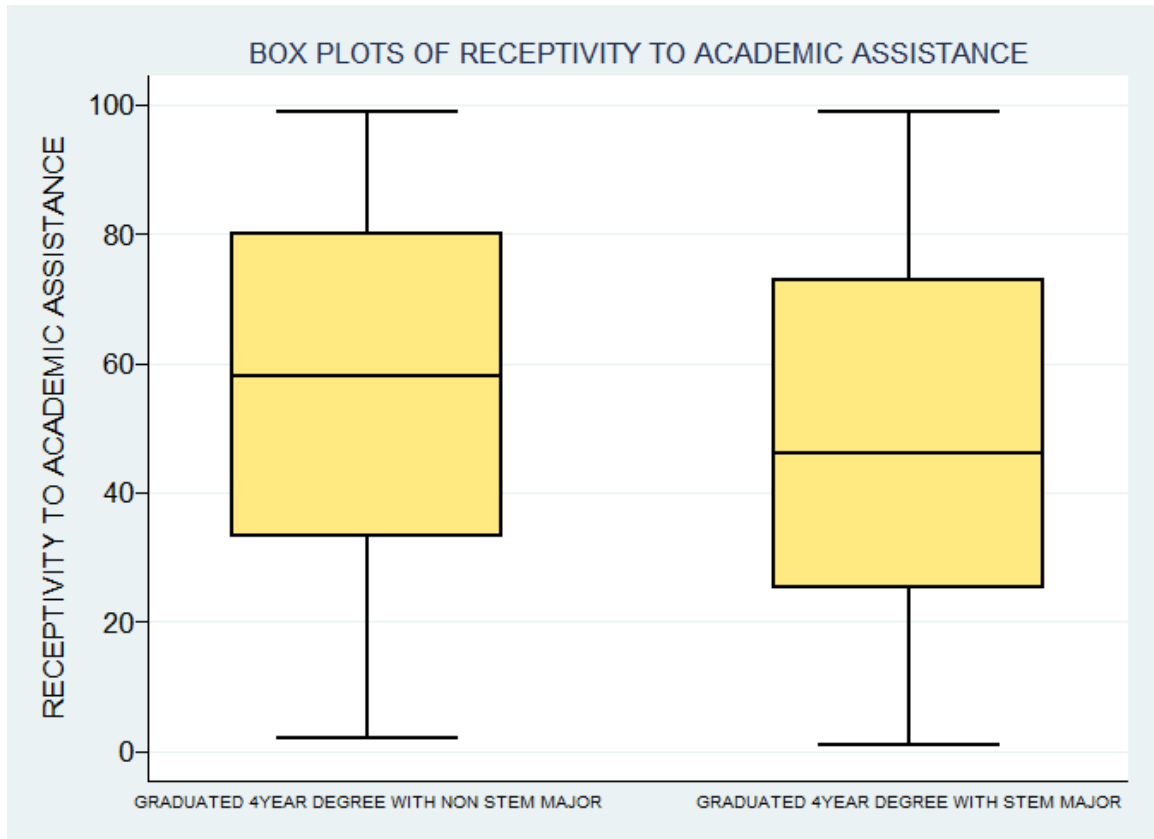
**Inference:** The median self-reported college prep of STEM graduates is 10 percentile greater than the NONSTEM students.



**FIGURE 52: BOX PLOT OF DESIRE TO FINISH PERCENTILE**

**Inference:** The median of Desire to finish percentile for NONSTEM major graduates is 65 and median of Desire to finish percentile for STEM major graduates is 59. The median NONSTEM graduates is 6 Percentile higher than STEM students.





**FIGURE 53: BOX PLOT OF RECEPTIVITY TO ACADEMIC ASSISTANCE**

Data are median with the third quartile (Q3) and first quartile (Q1).

The median receptivity to academic assistance for non-STEM graduates is 59 (80-33), and for STEM graduates is 46 (73-25).

**Inference:** The median receptivity to academic assistance of NONSTEM graduates is 13 percentile less than the STEM students.

This research now considers only those students who were retained in higher education, either at a 4 year or 2 year institution. Which factors impacted their decision to remain and eventually graduate with a STEM degree from a 4 year institution?

First step in the analysis is running a single logit model with each independent predictor variable. Table 2 below shows the variables that demonstrated the significance level of less than 0.25, the threshold level for entry into the initial logistic regression model. An initial logistic regression model is then presented. The model development continues similar to the previous model using the backwards elimination process to identify and remove the least significant independent variable and rebuild the model until only significant or impactful variables remain. The final regression model was constructed using all significant predictors. A threshold of 0.05 ( $\alpha = 0.05$ ) was used to identify all independent variables with significant association.

**TABLE 4: SINGLE LOGISTIC REGRESSION ANALYSIS OUTPUT BETWEEN STEM STUDENT GRADUATION AND INDEPENDENT VARIABLES**

<b>VARIABLE</b>	<b>COEFFICIENT</b>	<b>STD.ERROR</b>	<b>PVALUE</b>
Gender	-.3257721	.1802903	0.071
Transfer Percentile	-.0100385	.004009	0.012
Desire to Finish Percentile	-.0048492	.0031786	0.127
Academic Assistan	-.0087655	.0032348	0.007

Career Counseling	-.0044492	.0031986	0.164
Social Enrichment	-.0041866	.0034207	0.221
Family Emotional Support Percentile	.0043116	.0029603	0.145
Self-reported College Prepercentile	.0151966	.0036691	0.000
Math and Science Confidence Percentile	.0151607	.0038941	0.000
Career Closure Percentile	.0051642	.0031026	0.096
College Prep Percentile	.0148076	.0036625	0.000
Highest degree Sought Percentile	-.0045107	.0029862	0.131
High School Grade Percentile	.0178433	.0051716	0.001
Social Ability Percentile	-.0093047	.0030525	0.002
Desire to Transfer Percentile	-.0103407	.004026	0.010
MaxACTSAT Score	.1481677	.0263937	0.000
Distance From Campus	-.0008069	.0003914	0.039
HSGPA	.633918	.2558326	0.013

Class Percent	-.0175153	.0051053	0.001
PELL	-.2601014	.1881804	0.167
Major Code			
2	.778305	.5949523	0.191
3	.4079312	.4551796	0.370
4	-.184429	.4377278	0.674
5	.3364722	.6761234	0.619
6	.603535	.4554797	0.185
Degree Sought			
2	-.3162612	.2089235	0.130
3	-.2770775	.2408462	0.250
Fathers Education			
4	-.4911206	.2452528	0.045
5	.0152675	.2266483	0.946
6	-.315081	.2913716	0.280

Mothers Education	-.2238981	.2426815	0.356
4	.2722178	.2338496	0.244
5	-.0045353	.2883573	0.987
6			
Transfer Desire	-.4399139	.1917347	0.022
Coded Senior Year Grades			
2			
3	-.4904291	.1927184	0.011
	-.9209465	.3355627	0.006

## OUTPUT WINDOW 3: LOGISTIC REGRESSION OUTPUT WINDOW OF INITIAL REGRESSION MODEL WITH ALL 29 SIGNIFICANT FACTORS (STEM VS. NON-STEM)

```
. logit STEMANDNONSTEM i.CodifiedGender TransferPercentile DesiretoFinishPercentile ReceptivitytoAcademicAssistan ReceptivitytoCareerCouns
> eling ReceptivitytoSocialEnrichment FamilyEmotionalSupportpercen Selfreportedcollegepreperc MathandScienceConfidenceper CareerClosureperc
> entile collegeprepercentile Highestdegreesoughtpercentil Highschoolgradespercentile Sociabilitypercentile DesiretoTransferpercentile M
> axACTSATscore Distancefromcampus HSGPA Classpercent i.CodedPELL i.MAJORCODE i.DegreeSought i.FathersEducation i.MothersEducation i.Tra
> nsferDesire i.CodedSeniorYearGrades i.CodedWork
```

```
note: 3.DegreeSought omitted because of collinearity
Iteration 0: log likelihood = -318.49776
Iteration 1: log likelihood = -259.75265
Iteration 2: log likelihood = -259.22205
Iteration 3: log likelihood = -259.22121
Iteration 4: log likelihood = -259.22121
```

```
Logistic regression                                Number of obs      =           461
                                                    LR chi2(38)        =          118.55
                                                    Prob > chi2        =           0.0000
Log likelihood = -259.22121                        Pseudo R2         =           0.1861
```

STEMANDNONSTEM	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.CodifiedGender	.0219762	.2550011	0.09	0.931	-.4778167	.5217691
TransferPercentile	-.0362744	.056351	-0.64	0.520	-.1467203	.0741716
DesiretoFinishPercentile	-.0109892	.0049293	-2.23	0.026	-.0206505	-.001328
ReceptivitytoAcademicAssistan	-.0013025	.0055063	-0.24	0.813	-.0120947	.0094897
ReceptivitytoCareerCounseling	.0008534	.0054478	0.16	0.876	-.0098241	.0115308
ReceptivitytoSocialEnrichment	.0061692	.0057614	1.07	0.284	-.005123	.0174613
FamilyEmotionalSupportpercen	.0059521	.0039782	1.50	0.135	-.001845	.0137493
Selfreportedcollegepreperc	.0072563	.0329992	0.22	0.826	-.057421	.0719337
MathandScienceConfidenceper	.0094493	.0055131	1.71	0.087	-.0013562	.0202549
CareerClosurepercentile	.0099392	.0046535	2.14	0.033	.0008184	.01906
collegeprepercentile	.0003639	.0330706	0.01	0.991	-.0644532	.0651811
Highestdegreesoughtpercentil	-.0134609	.0079706	-1.69	0.091	-.0290831	.0021612
Highschoolgradespercentile	.0268107	.0417288	0.64	0.521	-.0549763	.1085976
Sociabilitypercentile	-.0101102	.0041367	-2.44	0.015	-.0182181	-.0020024
DesiretoTransferpercentile	.0242208	.0583325	0.42	0.678	-.0901089	.1385504
MaxACTSATscore	.1129146	.0409759	2.76	0.006	.0326034	.1932258
Distancefromcampus	-.0009426	.0005398	-1.75	0.081	-.0020006	.0001155
HSGPA	-.3859415	.3759011	-1.03	0.305	-1.122694	.3508112
Classpercent	-.0058918	.0085912	-0.69	0.493	-.0227302	.0109467
1.CodedPELL	-.3533643	.2505473	-1.41	0.158	-.844428	.1376994
MAJORCODE						
2	1.307224	.7065498	1.85	0.064	-.0775878	2.692037
3	1.300397	.5644571	2.30	0.021	.1940812	2.406712
4	.808499	.551449	1.47	0.143	-.2723212	1.889319
5	1.986137	.8641542	2.30	0.022	.292426	3.679848
6	2.073465	.5856439	3.54	0.000	.9256236	3.221306
DegreeSought						
2	.2678999	.4765254	0.56	0.574	-.6660727	1.201872
3	0	(omitted)				
FathersEducation						
4	-.9797	.3256132	-3.01	0.003	-1.61789	-.3415098
5	-.4210672	.320117	-1.32	0.188	-1.048485	.2063507
6	-1.053585	.4291144	-2.46	0.014	-1.894633	-.2125358
MothersEducation						
4	-.2494916	.3130593	-0.80	0.425	-.8630766	.3640934
5	.3499954	.3311413	1.06	0.291	-.2990295	.9990204
6	.432086	.417795	1.03	0.301	-.3867771	1.250949
1.TransferDesire	-.0086582	.5035166	-0.02	0.986	-.9955326	.9782162
CodedSeniorYearGrades						
2	.5128924	1.074713	0.48	0.633	-1.593506	2.61929
3	.8421641	2.418185	0.35	0.728	-3.897392	5.58172
CodedWork						
2	-.7470935	.3987248	-1.87	0.061	-1.52858	.0343926
3	-.4422284	.3341707	-1.32	0.186	-1.097191	.2127342
4	-.9590507	.3796904	-2.53	0.012	-1.70323	-.2148713
_cons	-3.36492	4.149602	-0.81	0.417	-11.49799	4.768149

Since the p-value is less than 0.05, the model itself is statistically significant. Here I dropped Degree Sought because of collinearity It is removed and regression ran again.

## OUTPUT WINDOW 4: LOGISTIC REGRESSION OUTPUT WINDOW WHEN DEGREE SOUGHT IS DROPPED (STEM VS. NON-STEM)

```
. logit STEMANDNONSTEM i.CodifiedGender TransferPercentile DesiretoFinishPercentile ReceptivitytoAcademicAssistan ReceptivitytoCareerCouns
> eling ReceptivitytoSocialEnrichment FamilyEmotionalSupportpercen Selfreportedcollegeprepper MathandScienceConfidenceper CareerClosureperc
> entile collegeprepperpercentile Highestdegreesoughtpercentil Highschoolgradespercentile Sociabilitypercentile DesiretoTransferpercentile M
> axACTSATscore Distancefromcampus HSGPA Classpercent i.CodedPELL i.MAJORCODE i.FathersEducation i.MothersEducation i.TransferDesire i.C
> odedSeniorYearGrades i.CodedWork
```

```
Iteration 0: log likelihood = -318.49776
Iteration 1: log likelihood = -259.87983
Iteration 2: log likelihood = -259.38034
Iteration 3: log likelihood = -259.37956
Iteration 4: log likelihood = -259.37956
```

```
Logistic regression               Number of obs   =           461
                                LR chi2(37)       =          118.24
                                Prob > chi2        =           0.0000
Log likelihood = -259.37956       Pseudo R2      =           0.1856
```

STEMANDNONSTEM	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.CodifiedGender	.0063413	.2536117	0.03	0.980	-.4907285	.5034111
TransferPercentile	-.0331922	.0561709	-0.59	0.555	-.1432851	.0769006
DesiretoFinishPercentile	-.0111496	.0049223	-2.27	0.024	-.0207972	-.001502
ReceptivitytoAcademicAssistan	-.0014407	.0054954	-0.26	0.793	-.0122116	.0093301
ReceptivitytoCareerCounseling	.0007516	.0054417	0.14	0.890	-.0099139	.0114171
ReceptivitytoSocialEnrichment	.0063568	.0057469	1.11	0.269	-.0049068	.0176205
FamilyEmotionalSupportpercen	.0060009	.0039764	1.51	0.131	-.0017926	.0137945
Selfreportedcollegeprepperperc	.0058366	.0326594	0.18	0.858	-.0581747	.0698479
MathandScienceConfidenceper	.0093034	.0055072	1.69	0.091	-.0014905	.0200973
CareerClosurepercentile	.0101489	.0046402	2.19	0.029	.0010543	.0192435
collegeprepperpercentile	.0019062	.0327174	0.06	0.954	-.0622188	.0660312
Highestdegreesoughtpercentil	-.0096975	.0043148	-2.25	0.025	-.0181544	-.0012407
Highschoolgradespercentile	.0252532	.041637	0.61	0.544	-.0563538	.1068602
Sociabilitypercentile	-.0102453	.0041279	-2.48	0.013	-.0183359	-.0021547
DesiretoTransferpercentile	.0203433	.0580077	0.35	0.726	-.0933497	.1340363
MaxACTSATscore	.1132576	.0409244	2.77	0.006	.0330473	.1934678
Distancefromcampus	-.0009734	.0005374	-1.81	0.070	-.0020267	.0000798
HSGPA	-.3822258	.3755977	-1.02	0.309	-1.118384	.3539322
Classpercent	-.0055699	.0085869	-0.65	0.517	-.0224	.0112602
1.CodedPELL	-.347926	.2503587	-1.39	0.165	-.83862	.1427681
MAJORCODE						
2	1.331202	.7044469	1.89	0.059	-.049489	2.711892
3	1.314299	.562033	2.34	0.019	.2127351	2.415864
4	.8474807	.5450201	1.55	0.120	-.2207391	1.9157
5	2.013422	.8620627	2.34	0.020	.3238103	3.703034
6	2.116192	.5792014	3.65	0.000	.9809779	3.251405
FathersEducation						
4	-.9758742	.3252699	-3.00	0.003	-1.613391	-.3383569
5	-.4149389	.3200978	-1.30	0.195	-1.042319	.2124414
6	-1.059139	.4287422	-2.47	0.013	-1.899458	-.2188195
MothersEducation						
4	-.2464059	.3126508	-0.79	0.431	-.8591902	.3663785
5	.3462936	.331067	1.05	0.296	-.3025859	.995173
6	.4359471	.4175303	1.04	0.296	-.3823972	1.254291
1.TransferDesire	.0232123	.500535	0.05	0.963	-.9578182	1.004243
CodedSeniorYearGrades						
2	.4869782	1.073883	0.45	0.650	-1.617793	2.591749
3	.7496246	2.41413	0.31	0.756	-3.981982	5.481232
CodedWork						
2	-.7232559	.3960015	-1.83	0.068	-1.499405	.0528928
3	-.4229942	.3321156	-1.27	0.203	-1.073929	.2279404
4	-.9452799	.3785885	-2.50	0.013	-1.6873	-.20326
_cons	-3.363843	4.146108	-0.81	0.417	-11.49007	4.762381

Based on p values the least significant variable is gender ( $p = 0.980$ ), and it is dropped in the next run.

## OUTPUT WINDOW 5: LOGISTIC REGRESSION OUTPUT WINDOW WHEN FAMILY EMOTIONAL SUPPORT IS DROPPED (STEM VS. NON-STEM)

```
. logit STEMANDNONSTEM1 TransferPercentile DesiretoFinishPercentile Selfreportedcollegeprepperc CareerClosurepercentile Highestdegrees
> oughtpercentil Sociabilitypercentile MaxACTSATscore Distancefromcampus HSGPA Classpercent i.MAJORCODE i.FathersEducation i.CodedW
> ork
```

```
Iteration 0: log likelihood = -320.37348
Iteration 1: log likelihood = -269.44862
Iteration 2: log likelihood = -268.9661
Iteration 3: log likelihood = -268.9646
Iteration 4: log likelihood = -268.9646
```

```
Logistic regression              Number of obs   =      464
                                LR chi2(21)      =     102.82
                                Prob > chi2       =     0.0000
Log likelihood = -268.9646      Pseudo R2      =     0.1605
```

STEMANDNONSTEM1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
TransferPercentile	-.012883	.0050316	-2.56	0.010	-.0227448	-.0030211
DesiretoFinishPercentile	-.0088026	.0042435	-2.07	0.038	-.0171198	-.0004855
Selfreportedcollegeprepperc	.0125296	.0054201	2.31	0.021	.0019065	.0231528
CareerClosurepercentile	.0103569	.0040857	2.53	0.011	.0023491	.0183647
Highestdegreesoughtpercentil	-.0098671	.0040748	-2.42	0.015	-.0178535	-.0018807
Sociabilitypercentile	-.0069448	.0036841	-1.89	0.059	-.0141654	.0002758
MaxACTSATscore	.1294563	.0378774	3.42	0.001	.0552179	.2036947
Distancefromcampus	-.0010225	.0004949	-2.07	0.039	-.0019926	-.0000524
HSGPA	-.1287561	.3565246	-0.36	0.718	-.8275315	.5700192
Classpercent	-.0068719	.0076719	-0.90	0.370	-.0219085	.0081647
MAJORCODE						
2	1.245713	.6728046	1.85	0.064	-.07296	2.564386
3	1.37311	.5421472	2.53	0.011	.3105213	2.435699
4	.8451061	.5300876	1.59	0.111	-.1938464	1.884059
5	2.099097	.8403062	2.50	0.012	.4521269	3.746067
6	2.089877	.5638691	3.71	0.000	.9847143	3.195041
FathersEducation						
4	-.7832141	.2884707	-2.72	0.007	-1.348606	-.217822
5	-.1586141	.2732614	-0.58	0.562	-.6941965	.3769683
6	-.5329424	.3497489	-1.52	0.128	-1.218438	.1525528
CodedWork						
2	-.610333	.380827	-1.60	0.109	-1.35674	.1360741
3	-.3463921	.3172285	-1.09	0.275	-.9681487	.2753644
4	-.7784194	.3562033	-2.19	0.029	-1.476565	-.0802738
_cons	-2.029892	1.695091	-1.20	0.231	-5.35221	1.292426

Final list of variables that appear in table : Transfer percentile, desire to finish percentile, self-reported college prep percentile, career closure percentile, highest degree sought percentile, sociability percentile, max ACT SAT score, distance from campus, HSGPA,



class percent, major code, fathers education and work are hat impact retention to graduation in a Stem DEGREE for first year students selecting a STEM degree at WT.

**TABLE 5: FINAL LIST OF FACTORS WITH SIGNIFICANT EFFECT ON STUDENT  
RETENTION AND GRADUATION IN HIGHER EDUCATIONS WITH STEM MAJORS.**

<b>VARIABLE</b>	<b>P VALUE</b>
Transfer percentile	0.010
Desire to finish percentile	0.038
Self-reported college prep percentile	0.021
Career closure percentile	0.011
Highest degree sought percentile	0.015
Sociability percentile	0.058
Max act sat percentile	0.001
Distance from campus	0.038
HSGPA	0.118
Class percent	0.370

Major	
2	0.084
3	0.011
4	0.111
5	0.012
6	0.000
Fathers education	
4	0.07
5	0.502
6	0.123
Work	
2	0.100
3	0.275
4	0.123

## Observations

- When I drop the Mothers education, the coefficients of fathers' education 5 & 6 changed 26% and 22%. Then I did the Log likelihood ratio test, I got the p value = 0.192960024. Here p value is not significant so, I decided to drop Mothers education.
- When I drop HSGPA the log likelihood ratio test of p value = 0.001228794, Here p values is significant, so I kept the HSGPA back in the model.
- When I drop the class percent the log likelihood ratio test of P value = 4.28761E-08, Here P value is significant. So I kept the class percent back in the model.

## Summary

Following predictors were shown to influence student graduation with STEM majors:

1. Transfer percentile
2. Desire to finish percentile
3. Self-reported college prep percentile
4. Career closure percentile
5. Highest degree sought percentile
6. Sociability percentiles
7. Max ACT SAT score

8. Distance from campus
9. HSGPA
10. class percent
11. Major code
12. Fathers' education
13. work

## CHAPTER V: CONCLUSIONS

The results of this study revealed important factors affecting students' success in 4-year degree program and community college degree programs followed by factors affecting their success in STEM majors.

For these analyses, two logistic regression models were constructed using the LOGIT procedure in Stata. A threshold of 0.25 ( $\alpha = 0.25$ ) is used to identify independent variables in a single predictor logistic model with significant influence. Then with all significant independent variables, logistic regression was performed with backwards elimination process to identify and remove the least significant independent variable, and this process was repeated until there was no non-significant independent variables. The final regression model was constructed using all significant predictors. A threshold of 0.05 ( $\alpha = 0.05$ ) was used to identify all independent variables with significant association. Following seven predictors were shown to influence student graduation:

1. Math and science confidence percentile
2. study habits percentile
3. HSGPA
4. Class percent
5. PELL
6. Mothers' education
7. Senior year grades.

The median of math and science confidence for 4 year graduates and community college graduates is 4 percentile greater than drop outs. 4-year degree students' median study habits percentile was at least 4-8 percentile greater than community college graduates or drop outs, Descriptive analysis of data demonstrated that median GPA percentile of drop out students is 3-7 percentile less than the GPA percentile of 4-year degree graduates and community college graduates. The median class percent of 4-year degree graduates students was at least 11-16 percentile less than community college graduates or drop outs. About 1/3 of 4-year graduates received PELL, while about 1/2 of the community college graduates and drop out were received PELL. 8-15% of mothers have elementary education and some high school. 1-10% of mothers have some college education. 0.8-12% of mothers have bachelor's degree. 2-7% of mothers have a master's and professional degree. Proportion of students with C or lower grade in senior year is greater in drop outs (32%) than 4-year degree graduates (8.7%) and community college graduates (18.8%).

The p-values for the model as well as the seven predictors are all less than 0.05, indicating a statistical significance. Final factors that were determined to affect undergraduate student graduation can be used to improve students' success in higher education institutions. Following predictors were shown to influence student graduation with STEM majors:

1. Transfer percentile

2. Desire to finish percentile
3. Self-reported college prep percentile
4. Career closure percentile
5. Highest degree sought percentile
6. Sociability percentiles
7. Max ACT SAT score
8. Distance from campus
9. HSGPA, class percent
10. Major code
11. Fathers' education and work

The median of Transfer percentile for NONSTEM major graduates is 60 and median of Transfer percentile for STEM major graduates is 50. The median NONSTEM graduate is 10 Percentile higher than STEM students. The median of Desire to finish percentile for NONSTEM major graduates is 65 and median of Desire to finish percentile for STEM major graduates is 59. The median NONSTEM graduate is 6 Percentile higher than STEM students. The median of Career closure percentile for STEM major graduates is 53 and median of Career closure percentile for STEM major graduates is 83. The median STEM graduate is 30 Percentile higher than



NONSTEM students. The median sociability percentile of non-STEM graduates is 10 percentile greater than the STEM students. The median max ACT & SAT scores of non-STEM graduates is 2 percentile less than that of STEM graduates. The median of Distance from campus for STEM major graduates is 83 and median of Distance from campus for STEM major graduates is 67. The median STEM graduate is 16 Percentile higher than NONSTEM students. Median high school GPA of STEM graduates is 0.1 percentile greater than that of non-STEM graduates. The median class percent of non-STEM graduates is 7 percentile less than the STEM graduates. Almost about same percent of STEM and NON STEM graduates chose their majors. There is a small difference between the level of parent's education of STEM and non-STEM graduates. Less proportion of STEM graduates (19.5%) worked for 21-30 hours per week compared to non-STEM graduates (27.8%).

The p-values for the final model is less than 0.05, indicating a statistical significance. Final factors that were determined to affect student graduation can be used to increase students with STEM majors.

## REFERENCES

1. Sagenmuller, I. *MAIN FACTORS THAT AFFECT STUDENT RETENTION IN HIGHER EDUCATION*. 2016 [cited 2017 08/22/2017]; Available from: <https://www.u-planner.com/blog/main-factors-that-affect-student-retention-in-higher-education>.
2. *What is a STEM degree*. 2017 [cited 2017 08/22/2017]; Available from: <http://www.bestvalueschools.com/faq/what-is-a-stem-degree>.
3. *Science, Technology, Engineering and Math: Education for Global Leadership*. 2017.
4. *STEM Performance*. 2017.