**Module 7 Portfolio Milestone**

**Joseph Yevara**

**Colorado State University Global**

# MIS581-1: Capstone: Business Intelligence and Data Analytics

# Dr. Osama Morad

**02/04/2024**

**Abstract**

Higher education faces an ever-increasing desire to retain as many students as possible considering the cost of college and the decreasing amount of college eligible students. With this, many colleges and universities have started to look at factors leading towards increased student persistence rather than the broad student retention metric. Prevailing research on the topic suggests that student persistence is reliant on pre-entry attributes, goal commitments, effort, and institutional experiences. This project aims to look at different socioeconomic, academic, and economic factors and their impact on student persistence using data collected from an institution between the 2008/2009 and 2018/2019 academic years. Using logistic regression, the project has found which factors hold more predictive power that students will persist. Findings are somewhat consistent with previous research however, there are a few conclusions that were surprising such as scholarship holders being less likely to persist than non-scholarship holders. Future research into student persistence should take the findings of this project and further explore specific factors as this is a scattershot of different factors.

**Introduction**

Many colleges and universities are facing a growing concern surrounding enrollment as declines in the birth rate over the past couple decades are starting to affect the number of college eligible individuals graduating high school. This is often referred to as the “enrollment cliff” and is expected to see an 11% reduction in students at regional bachelor’s institutions by 2029 (Kline, 2019). Considering that student tuition and fees accounted for an average of 42% of total revenue at institutions, this could have major implications for the higher education landscape in the United States (Huiskes, 2023). In fact, reductions in enrollment and funding have accounted for most of the 15 non-profit colleges and university closures in 2023 (Moody, 2023).

Colleges and universities cannot change declining birth rates in the United States so they must focus on enrolling as many students as possible, and perhaps more importantly, retaining as many students as possible. Therefore, institutions pay careful attention to important metrics such as retention rates and student persistence. Steele (n.d.), defines retention rates as “the number of students that re-enroll from one year to the next,” and student persistence as an attempt to quantify “the drive, determination, and activities that propel students forward toward their degrees.” While retention rates are good to inform macro level decisions like budgets and salaries, student persistence informs micro level decisions to increase student success and the likelihood that individual students will be retained. For example, seeing that many commuter students are dropping out of an institution may inform said institution to create initiatives aimed at improving the commuter student experience.

## Purpose Statement

This capstone project is aimed at determining individual and larger economic factors that play pivotal roles within student persistence. Without a good understanding of these factors, institutions are forced to look at retention rates and make broader generalizations in an attempt to increase these metrics. The data set being utilized contains various undergraduate student information from before and during their time within the college as well as information surrounding the greater economic climate at the time of enrollment. Ideally, figuring out which factors had the most influence on whether or not these students dropped out or graduated will provide good direction for further research into those areas. For example, if certain majors are significantly more likely to drop out then others, further research could be conducted to understand why these trends occur in the first place. At the same time, understanding larger economic factors (country GDP, unemployment rate, etc.) and their predictive power over dropout rates could lend guidance to institutions on what to expect and account for these changes during certain economic climates.

**Aims & Objectives**

There are three major aims of the project which are to determine which socioeconomic, academic, and economic factors most affect student persistence. Socioeconomic factors are defined as variables attributed to individual students beyond academics, while academic factors are defined as variables attributed to individual student classroom performance. Examples of socioeconomic factors include parental education levels and financial aid while examples of academic factors include semester credits attempted/completed and major or study. Economic factors are defined as variables that can be attributed to every student at a given time as they relate to the larger economy of the country in which the university is located. Examples of economic factors include unemployment rate, inflation rate, and GDP.

Each aim has the same consequent objectives which start with understanding the variables contained within the data set. Getting a good grasp on the variables contained for socioeconomic and academic factors will help to further define what the student population looks like at the institution while economic factors will help to understand what the economic climates were when this data was collected. From there, the next objective is to analyze possible correlations between the variables outside of correlation with whether or not each student was retained. Any correlations found between the variables could lead to recommendations for further research as they may point to information outside the scope of this particular research project. This will also help to narrow down the variables for the next objective which is to quantify which variables hold the greatest predictive power over whether or not a student will persist at the university or dropout. As stated previously, this is the ultimate goal of the research project and leads to the final objective which is to provide recommendations based upon these results. Those recommendations can also draw upon research from other fields or national organizations into best practices in student persistence.

**Overview of Study**

This study is being conducted utilizing a pre-existing data set containing anonymous information from students enrolled at a university in Portugal from the 2008/2009 to 2018/2019 academic years. Within this data set are variables kept on the previously mentioned socioeconomic, academic, and economic factors. Each row of data refers to one student that was enrolled at the university in that time period and contains whether or not the student had graduated, dropped out, or was still enrolled when the data was collected. A set of descriptive and predictive analytics will be conducted using this data to find indicators of student persistence.

**Research Questions and Hypotheses**

Overall, the research question for this project is what factors lead to increased student persistence? However, this is extremely broad when reviewing the sheer amount of potential factors that might influence student persistence. Therefore there are three areas of focus and thus three research questions for this particular project:

1. Which socioeconomic factors increase student persistence?
2. Which academic factors increase student persistence?
3. Which economic factors increase student persistence?

**Hypotheses**

Each of the three research questions, listed above, have a hypothesis for this research project. All three hypotheses are fairly similar in nature but again, there are categories of data included within the dataset that must be given equal consideration in regards to student persistence so that the results of the project are easier to comprehend for these institutions.

1. Which socioeconomic factors increase student persistence?

* H1: There are no socioeconomic factors that create a statistically significant increase in whether or not students stay at the institution
* H2: There are socioeconomic factors that create a statistically significant increase in whether or not students stay at the institution

Socioeconomic factors within the dataset include demographic information, parental degrees/occupation, and student loans. These factors represent an opportunity for institutions to provide increased support to students closer to the time of enrollment so that they are better prepared for college.

1. Which academic factors increase student persistence?

* H1: There are no academic factors that create a statistically significant increase in whether or not students stay at the institution
* H2: There are academic factors that create a statistically significant increase in whether or not students stay at the institution

Academic factors within the dataset include course majors, grades received, and credits taken. These factors can show institutions where extra student support may be needed in certain academic pursuits and at certain milestones within a students’ college experience (i.e. after first semester, once grades fall below a certain threshold, etc.).

1. Which economic factors increase student persistence?

* H1: There are no economic factors that create a statistically significant increase in whether or not students stay at the institution
* H2: There are economic factors that create a statistically significant increase in whether or not students stay at the institution

Economic factors within the dataset include unemployment rate, gross domestic product, and inflation rate. These factors could give institutions an idea of what retention is going to look like in certain economic climates, allowing them to plan accordingly.

**Literature Review**

Steele & Douglas (2021) define student persistence “as the degree to which a student remains enrolled in their current degree” (pg. 1415). Colleges and universities are increasingly concerned with this metric not only because it represents expected revenue, but also because it has been used as a way to rank institutions of higher education (Gabi & Sharpe, 2021). Existing research has relied heavily on Tinto’s Interactionalist Theory of Student Departure in which Tinto argues that student persistence comes from pre-entry attributes, goal commitments, effort, and institutional experiences (Steele & Douglas, 2021). Although this reliance has also come with disclaimers that Tinto’s research relied heavily on mostly residential colleges and non-minority student populations (Steele & Douglas, 2021). This research project aims to provide more context on pre-entry attributes (socioeconomic factors), goal commitments, and effort (both academic factors). It is important to note however that this research project does not provide more context on institutional experiences which would be a good place for future research considering some research indicates that high-level engagement leads to higher student persistence even with lower academic performance (Hu, 2010).

**Socioeconomic Factors**

Existing research indicates that socioeconomic factors play a pivotal role in student persistence (Reynolds & Cruise, 2020). For example, socioeconomic status and parental education levels were found to influence student persistence with higher socioeconomic status and parental education levels leading to higher rates of student persistence (Reynolds & Cruise, 2020). This is especially true in more specialized majors such as STEM programs (Pedraza & Chen, 2022). Furthermore, transnational students (i.e. students from different countries) are shown to have lower levels of commitment, integration, and satisfaction (Steele & Douglas, 2021). These results shine importance on giving support to first generation students as they navigate the collegiate experience (Reynolds & Cruise, 2020) as well as transnational students (Steele & Douglas, 2021).

**Academic Factors**

Past research shows that there is a level of nuance when discussing the effect of academic factors on student persistence. On one hand there are studies that show that academic success within the first year has tremendous implications for student success however, at the same time, results are different when factoring in intersectionality of student identity and student’s major of study (Pedraza & Chen, 2022). Studies on academic success and student persistence also indicate the need for more faculty and tutoring support (Gabi & Sharpe, 2021). Academic success seems to be tied into students' sense of belonging in college which is why students do not reach out for support when they are struggling academically (Gabi & Sharpe, 2021). Struggling academically seems to create fear within students that they do not belong at the university and getting help reinforces this idea within their own minds (Gabi & Sharpe, 2021).

**Economic Factors**

Beyond socioeconomic status, there does not seem to be as much research done concerning greater economic factors and the role they play with student persistence. Research however, has been done in the area of greater economic factors and student retention (Garrett, 2022). This research suggests that while in the past economic downturn has led to an increase in higher education enrollment, that may still not be the case in modern times (Garrett, 2022). The 2008 recession in the United States saw an enrollment increase of 16% but the rising cost of higher education mixed with post-pandemic increases in mental health have seen less enrollment increases and lower student persistence numbers (Recchi, 2022). Recchi (2022), argues that institutions of higher education do not provide adequate support for students to see such an increase as the 2008 recession. This support will need to not only be in the form of financial aid but also in student wellness programs as the cost continues to rise in the years ahead.

**Research Design**

**Methodology**

This is a quantitative research project as all data collected is either nominal or ordinal. Each variable collected from students has been categorized as seen within the data dictionary in Figure 1. There is no qualitative data present in the data collected which could be a direction for further research to better understand why students persisted at the institution.

**Figure 1**

*Data dictionary for dataset*

| **Variable** | **Type** | **Description** |
| --- | --- | --- |
| Marital Status | Numeric | Marital status of student |
| Application Mode | Numeric | Method of application to the university |
| Application Order | Numeric | Order in which application to the university was completed |
| Course | Numeric | Major student is pursuing |
| Daytime/evening attendance | Numeric | Daytime or evening attendance |
| Previous qualification | Numeric | Previous educational qualifications of student prior to enrolling |
| Nationality | Numeric | Nationality of student |
| Mother's qualification | Numeric | Previous educational qualifications of student's mother prior to enrolling |
| Father's qualification | Numeric | Previous educational qualifications of student's father prior to enrolling |
| Mother's occupation | Numeric | Occupation of student's mother |
| Father's occupation | Numeric | Occupation of student's father |
| Displaced | Numeric | Whether or not a student comes from the area in which the institution resides |
| Educational special needs | Numeric | Does the student need special accommodations for their education? |
| Debtor | Numeric | Has the student taken loans to finance their educational pursuit? |
| Tuition fees up to date | Numeric | Are the student's tuition fees up to date? |
| Gender | Numeric | Gender of student |
| Scholarship holder | Numeric | Does the student hold a scholarship? |
| Age at enrollment | Numeric | Age of student at the time of enrollment |
| International | Numeric | Is the student from a foreign country? |
| Curricular units 1st sem (credit) | Numeric | How many credits student had completed by the beginning of first semester |
| Curricular units 1st sem (enroll) | Numeric | How many credits were attempted in first semester of enrollment |
| Curricular units 1st sem (evalua) | Numeric | How many credits were evaluated in first semester of enrollment |
| Curricular units 1st sem (approv) | Numeric | Amount of credits student was given after completion of first semester of enrollment |
| Curricular units 1st sem (grade) | Numeric | Average grade received by student after first semester of enrollment |
| Curricular units 1st sem (withou) | Numeric | How many credits were not evaluated in first semester of enrollment |
| Curricular units 2nd sem (credit) | Numeric | How many credits student had completed by the beginning of second semester |
| Curricular units 2nd sem (enroll) | Numeric | How many credits were attempted in second semester of enrollment |
| Curricular units 2nd sem (evalua) | Numeric | How many credits were evaluated in second semester of enrollment |
| Curricular units 2nd sem (approv) | Numeric | Amount of credits student was given after completion of second semester of enrollment |
| Curricular units 2nd sem (grade) | Numeric | Average grade received by student after second semester of enrollment |
| Curricular units 2nd sem (withou) | Numeric | How many credits were not evaluated in second semester of enrollment |
| Unemployment rate | Numeric | Portuguese unemployment rate at time of enrollment |
| Inflation rate | Numeric | Portuguese inflation rate at time of enrollment |
| GDP | Numeric | Portuguese Gross Domestic Product at time of enrollment |
| Target | Categorical | Whether student graduated, dropped out, or is still enrolled |

**Methods**

Each hypothesis will go through a similar quantitative methodology. The only difference will be at the beginning of the project in which all variables within the data set are categorized into one of the three kinds of factors. Afterwards, the first step is to conduct descriptive statistics tests to get a good understanding of the data. These tests will be a mix of summary statistics and frequency distribution which will provide insight as to what is the general student population and economic state of the university and its country. Correlation analysis will follow these tests to ensure that none of the variables are highly correlated with each other. This is done to minimize the risk of multicollinearity in the forthcoming regression analysis. Any collinearity would damage interpretation of the final models as we could not definitively conclude that the results are based upon the factors themselves or other factors contained within the model (Saslow, 2018). Finally, a logistic regression analysis will be conducted for each category to understand which variables represent a positive relationship towards graduation and continued enrollment at the institution. From there conclusions can be made about which factors increase student persistence.

**Limitations**

The most obvious limitation to this study is its applicability to higher education institutions outside of Portugal. However, previous literature does seem to draw many parallels between factors influencing student persistence in European universities versus universities in other parts of the world (Gabi & Sharpe, 2021). Still, there is a need to do further research in other countries that choose to interpret the results of this study. Furthermore, the data collected for this study was done prior to the COVID-19 pandemic which Garrett (2022) mentions has changed the economic landscape of universities and subsequent enrollment. Therefore, conclusions on economic factors should be viewed with some scrutiny and possibly checked using data from after the COVID-19 pandemic.

**Ethical Considerations**

There are a few major ethical considerations presented with this type of data. Arguably the most important is the use of the demographic data contained within the data set. Notably, one variable contained within the data set is nationality. Call et. al (2022), argue that all demographic data must be carefully collected and analyzed to ensure ethical inferences and conclusions are drawn from said data. For example, if nationality is a factor that influences student persistence, careful reporting must be done to ensure that this does not end up creating policies that oppress students from different backgrounds. The intent should be to find strategies that make it easier for students to graduate from colleges and universities and not preclude them from these opportunities.

It is also important to note that any conclusions from this research project only show colleges and universities factors that will need further research conducted. Just because this project may show that students in a certain major do not persist at a high rate does not mean that major should be cut. Instead, it means that the institution should do further research into that particular major and find ways to create more support for those students. In other words, it is unethical for any institution to make massive changes based solely upon this project without conducting further research into the factors that are noted at its conclusion.

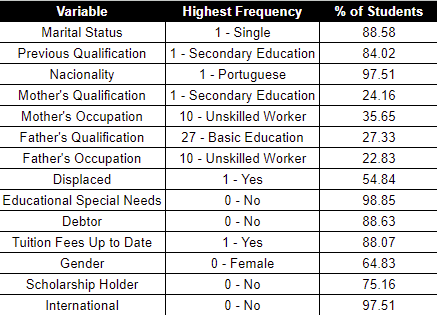
**Findings**

**Socioeconomic Factors**

Figure 1 shows the highest frequency for each individual factor included in the socioeconomic factors. An interesting finding is that most students at the institution have parents who are unskilled workers however, mother’s were more likely to hold secondary education while father’s were more likely to hold only a basic level of education. This may have been a cultural difference in Portugal versus western nations. Furthermore, the average age of enrollment for students was 23 years old.

**Figure 1**

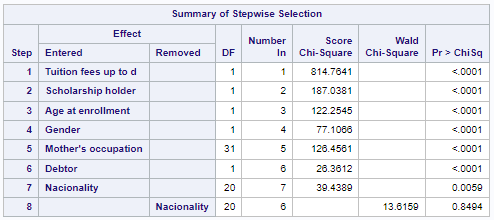
*Highest frequency for SES factors*

****

A stepwise selection method was used for the logistic regression which showed that previous qualification was not a significant predictor for graduation so that was taken out of the model. Figure 2 shows the stepwise selection method without previous qualification as an included variable. Figure 3 shows the significant variables that were a result of this second logistic regression model. Occupations for mothers that were 10, 32, and 7 (Unskilled Workers, Personal Care Workers, and Farmers) had students that were more likely to persist towards graduation. Likewise, male students that were in debt are also more likely to persist towards graduation. On the other hand, having tuition fees up to date and holding a scholarship meant the student was less likely to graduate. Furthermore, younger students were more likely to graduate at the institution. All of this information combined means that we can reject the null hypothesis and conclude that there are socioeconomic factors that increase the likelihood that a student will persist at the university.

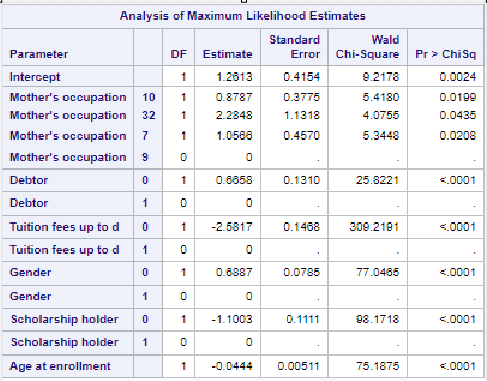
**Figure 2**

*Summary of stepwise selection for logistic regression*

****

**Figure 3**

*Significant predictors of graduation from logistic regression*

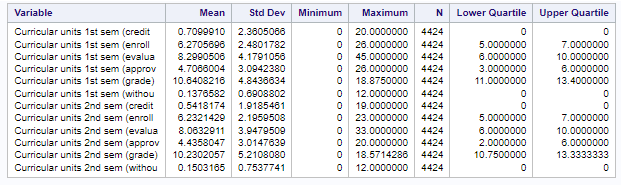
****

**Academic Factors**

Figure 4 shows the summary statistics for curricular units data presented in the data set. Of note, the institution seems to have a student population that sees slight declines in grades and completed credit hours from year one to year two. Figure 5 shows the distribution of course types amongst the data. Courses labeled “12” and “9” are the highest enrolled courses at the university and refer to nursing and management students.

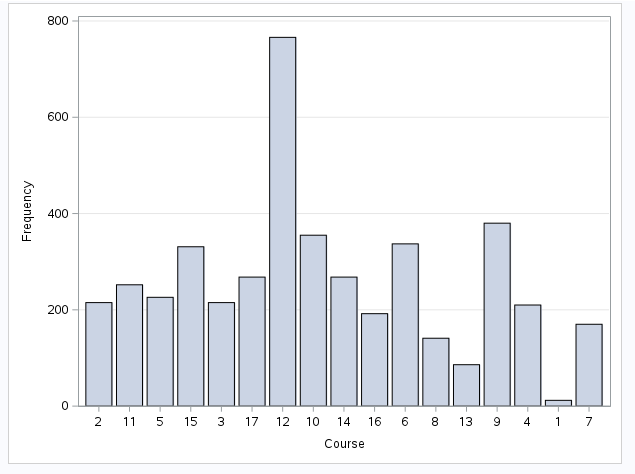
**Figure 4**

*Summary statistics for credit variables in academic factors*



**Figure 5**

*Distribution of courses*

****

Unfortunately, there were many variables that were closely correlated with each other in this factor group. Most of these variables contained a Pearson correlation coefficient of 0.5 or higher. Therefore, the decision was made to incorporate two non-correlated variables from these into the logistic regression model; 2nd semester credits enrolled and 2nd semester grades. These two variables are highly correlated (Pearson Correlation Coefficient) with their 1st semester counterparts as shown in Figure 6 therefore, it seemed ok to use the 2nd semester as an indication of 1st semester results.

**Figure 6**

*Correlation analysis of academic variables*

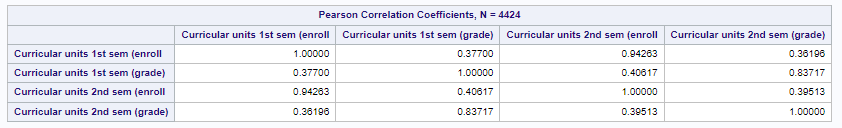
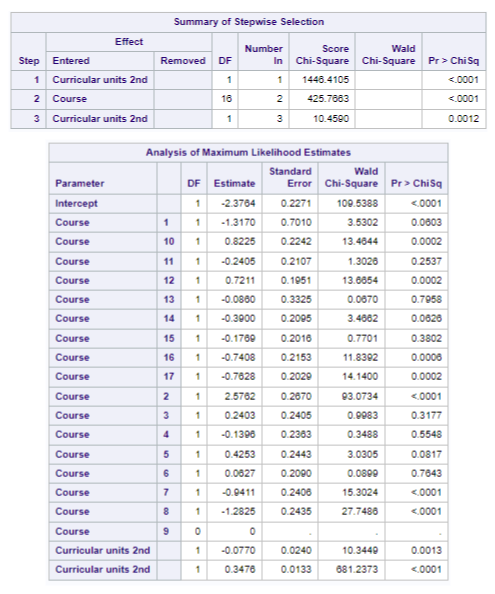
**

Figure 7 shows the logistic regression output for academic factors. Significance was only shown for the included 2nd semester variables and course variable. Unsurprisingly, 2nd semester grades showed that better grades indicates a higher probability for staying at the university. Furthermore, taking less credits for 2nd semester showed a higher probability of staying at the university.

Of the courses that were significant indicators in the model, there were a few that showed higher probability of graduation and some that showed lower probability. For instance, 10, 12, and 2 (Social Service, Nursing, and Animation) all showed a higher probability of graduation while 16, 17, 7, and 8 (Education, Management (evening), Engineering, and Equiniculture) all showed a lower probability. All of these results let us conclude by rejecting the null hypothesis and stating that their are academic factors that increase student persistence at the university.

**Figure 7**

*Logistic regression output for academic factors*

****

**Economic Factors**

Figure 8 shows descriptive statistics for the three variables selected for economic factors. There are no outliers for each of the variables and there is a fairly large range for each variable which is good for this project as it means this should give a pretty good understanding of the effects on student persistence. Unemployment rate averaged 11.6%, inflation rate averaged 1.2% and GDP averaged 0.002.

**Figure 8**

*Summary statistics & boxplots of economic factor variables*

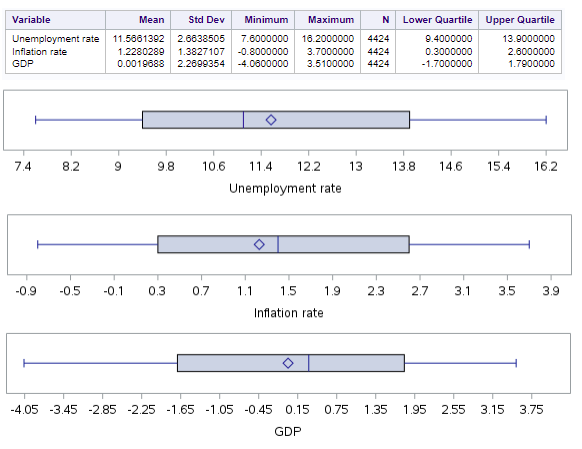


Figure 9 shows the correlation analysis done between the economic variables. None of the variables have a Pearson correlation coefficient above 0.5 and therefore are not highly correlated. This means that all variables can be used for the logistic regression test without fear of multicollinearity.

**Figure 9**

*Correlation tests for economic factors*

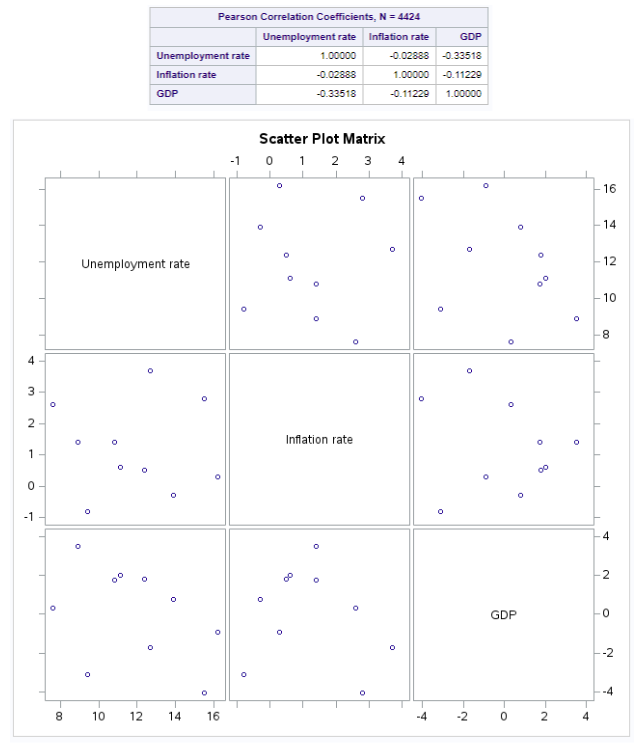
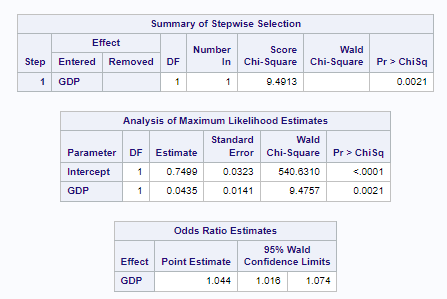


Figure 10 shows the results of the logistic regression test. A stepwise selection method was used which only included GDP as a predictor variable as the other variables did not meet significance requirements. The odds ratio shows a positive correlation with the graduation meaning that as GDP increases, so does the likelihood that the student graduates or stays enrolled at the institution. This means that we reject the null hypothesis and conclude that there are economic factors that increase the likelihood of a student to stay at the institution.

**Figure 10**

*Logistic regression output for economic factors*



**Conclusion**

Based on the findings of this research project, there are a multitude of different factors that can increase the likelihood of student persistence in higher education. Granted, some of these factors are surprising given previous research including that scholarship holders are less likely to persist. Either way, this project should give good direction into further research on the matter by lessening the amount of factors that need to be studied. As mentioned previously, each group of factors researched give guidance to universities in different ways. Socioeconomic factors allow universities to support students before they even start taking classes. Academic factors allow universities to support students as soon as grades or enrollment have been completed and economic factors allow universities to anticipate support needed by their students in different economic climates.

**Recommendations**

It is recommended to do further research into the areas that increase student persistence. For socioeconomic factors, doing more qualitative research into the effect of a student's mother’s occupations would be interesting and may yield some good insight into what motivates students from these backgrounds to finish their education. Similar reasons would apply for further research into why male students and debtors are more likely to persist. For academic factors, further research should be done into the differences between courses that have higher probability of graduation versus those that have lower probabilities. This seems like a good way to share knowledge between disciplines to benefit future students. Finally, for economic factors, further research can be done into how economic realities influence student persistence.

**References**

Call, C. C., Eckstrand, K. L., Kasparek, S. W., Boness, C. L., Blatt, L., Jamal-Orozco, N., Novacek, D. M., & Foti, D. (2022). An Ethics and Social Justice Approach to Collecting and Using Demographic Data for Psychological Researchers. *Perspectives on Psychological Science*.

Gabi, J., & Sharpe, S. (2021). Against the odds: an investigation into student persistence in UK higher education. *Studies in Higher Education*, *46*(2), 198–214. <https://doi.org/10.1080/03075079.2019.1626364>

Garrett, R. (2022, August 2). *Inflation nation: The latest data on college enrollment in today’s on-edge economy*. Encoura. <https://encoura.org/inflation-nation-the-latest-data-on-college-enrollment-in-todays-on-edge-economy/>

Hu, S. (2011). Reconsidering the Relationship Between Student Engagement and Persistence in College. *Innovative Higher Education*, *36*(2), 97–106. <https://doi.org/10.1007/s10755-010-9158-4>

Huiskes, H. (2023, May 25). *State Support for Higher Ed Continues to Rise. Yet Public Colleges Still Face Headwinds.* The Chronicle of Higher Education; The Chronicle of Higher Education. <https://www.chronicle.com/article/state-support-for-higher-ed-continues-to-rise-yet-public-colleges-still-face-headwinds>

Kline, M. (2019). *The Looming Higher Ed Enrollment Cliff*. CUPA-HR. <https://www.cupahr.org/issue/feature/higher-ed-enrollment-cliff/>

Moody, J. (2023, December 21). *A Look Back at College Closures and Mergers*. Inside Higher Ed. <https://www.insidehighered.com/news/business/financial-health/2023/12/21/look-back-college-closures-and-mergers-2023>

Pedraza, L., & Chen, R. (2022). Examining Motivator Factors of STEM Undergraduate Persistence through Two-Factor Theory. *Journal of Higher Education*, *93*(4), 532–558. <https://doi.org/10.1080/00221546.2021.1999722>

Recchi, C. (2022, August 10). *Addressing student persistence on the precipice of an economic downturn*. EdSights. <https://www.edsights.io/post/addressing-student-persistence-on-the-precipice-of-an-economic-downturn>

Reynolds, J., & Cruise, S. (2020). Factors that Influence Persistence Among Undergraduate Students: An Analysis of the Impact of Socioeconomic Status and First-Generation Students. *Interchange, 51*(2), 199-206. <https://doi.org/10.1007/s10780-020-09408-y>

Steele, A. R., & Douglas, H. E. (2021). Investigating the educational experiences of transnational students: Differences in academic integration, social integration, and institutional and goal commitment. *British Journal of Educational Psychology*, *91*(4), 1414–1433. <https://doi.org/10.1111/bjep.12424>

Steele, J. (n.d.). *Student Persistence vs Student Retention: What Is The Difference?* Suitable. Retrieved December 22, 2023, from <https://www.suitable.co/knowledge-center/blog/student-persistence-vs-student-retention-what-is-the-difference>