University of California, Los Angeles

Winter 2025

Statistics 140XP Final Project

Analyzing Academic Performance in Statistics Undergraduates: Course Combinations and Transfer Student Achievement

By Hannah Steinberg, Gabriel Pham, Peiyuan Lee,
Mari Yamamoto, Jasmine Yung, Bella Gordon

Table of Contents

Introduction

Academic success in undergraduate education can be influenced by many external factors, including course selection and student background. While a field like statistics may seem relatively structured, understanding the nuances between required curriculum and course combinations can help students make better informed decisions, optimizing learning to the individual.

Students also have various academic backgrounds and experiences, and this can also affect how they interact with the academic environment. Transfer students who attended a community college or other university will approach upper-division courses and education with a different perspective compared to those that will have attended the same university for all four years. In other words, the barriers of academic success are exceedingly different for students who attended the university starting as a first year, and those who transferred[1].

Some of the factors that influence transfer students' performance can be their previous accomplishments. In a study focusing on the strongest predictors of academic success (measured via GPA) for transfer students entering into a four-year university, it was found that the most significant predictor was the students' GPA earned at the community college, with a higher community college GPA correlating to a higher university GPA[4]. Transfer students accepted into UCLA for the 2022-2023 and 2023-2024 academic years tended to have near-perfect GPAs, with the median being 3.90 and 3.88 out of 4.00, respectively (and with upper quartiles of 4.00 both years). This also indicates that it can be expected that transfer students would demonstrate a similar high academic success level within this report[7,8]. Similarly, high school GPA was found to be the strongest predictor of a four year student's success in university; applying this to our specific population, four year UCLA students accepted into UCLA for the 2022-2023 and 2023-2024 academic years had impressive high school GPAs[2,5,6]. The interquartile unweighted GPAs ranges for both years were 3.95-4.00 with the medians being 4.00, indicating that it could be expected that four year students, correspondingly, succeed academically[5,6]. With both groups having similar strong performances upon entry into UCLA and predicted results, there is intrigue into whether the likeness between the two populations continues while at UCLA or a divergence occurs.

Overall, this study aims to analyze these factors and provide insights into potential distinctions between student experiences. These findings may contribute to a deeper understanding of academic performance in Statistics & Data Science at UCLA, with practical applications in designing statistics curriculum and improving academic success.

Research Questions

To better understand academic performance among Statistics majors at UCLA, this study explores the following questions:

(1) Which combination of Statistics and Math courses results in the highest overall GPA among Statistics majors at UCLA?

(2) Is there a difference between transfer students and four-year UCLA students in GPA for upper division statistics courses?

Methods

Data Source

The data for this study is sourced from UCLA's institutional database, which contains information on undergraduate students at UCLA. The dataset includes student demographics, course enrollments, when they have enrolled in these courses, their academic majors, and grades received. For this report, we will be focusing on students enrolled in the Statistics & Data Science major.

Data Cleaning & Preprocessing

Data preprocessing and exploratory analysis were conducted using Python and R. We begin by cleaning the data to ensure accuracy and relevance; checking for duplicate records, handling missing values, and standardizing categorical variables. We then filter the dataset to include only students who have declared Statistics & Data Science as their major, ensuring that our analysis remains focused on the target population. We also filter by course, focusing only on students enrolled in upper division and elective courses that contribute to the Statistic & Data Science major. Additionally, we encode categorical variables where necessary and transform grade data into a numerical GPA scale for consistency in analysis, focusing on those enrolled in upper-division courses and classes that contribute to the major. This process ensures that the analysis is both meaningful and aligned with the academic structure of the program.

The courses that were included in the cleaned dataset are shown in Table 1.

Major Requirements	**Elective Courses**
STATS 100A, STATS 100B, STATS 100C, STATS 101A, STATS 101B, STATS 101C, STATS 102A, STATS 102B, STATS 102C, STATS 140XP (IP grade), STATS 141XP	STATS 112-199, MATH 115A, MATH 131A, MATH 131B, MATH 151A, MATH 151B, MATH 156, MATH 170B, MATH 171, MATH 177, MATH 178A, MATH 178B, MATH 178C, COMM 153 or 188C, DGT HUM 101, ECON 143, ECON 144, ECON 147, EE BIOL C172, PSYCH 142H

Table 1. Statistics & Data Science Major Courses

Exploratory Data Analysis

Initial Count Comparison

To gain initial insights into the dataset, we conducted exploratory data analysis by examining student enrollment patterns based on admission type and course load. We looked at the distribution of freshman vs. transfer admit counts (Figure 1). We also looked at the number of courses taken per student, identifying variations in course load (Figure 2). This analysis helped us understand enrollment trends and establish a foundation for further investigation into academic performance and course-taking behavior.

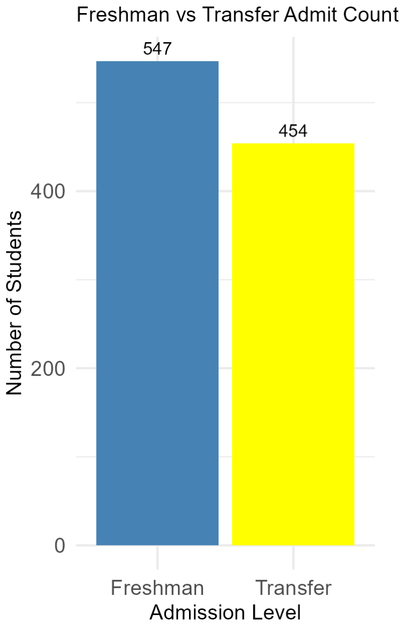

Figure 1. Statistics Majors Freshman vs.
Transfer Admit Count

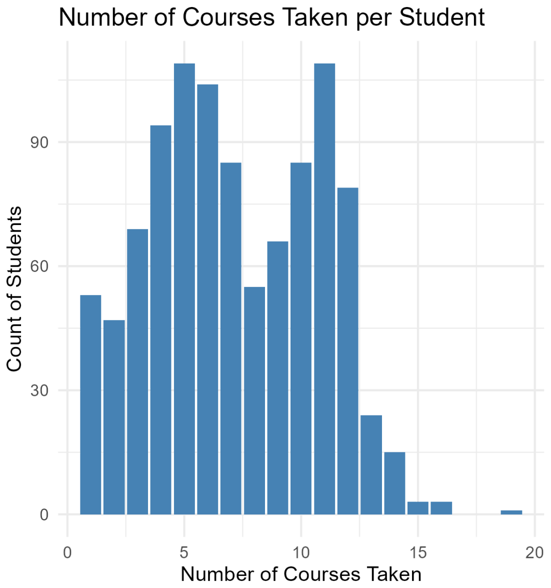

Figure 2. Number of Statistics Courses Taken Per Student

Tableau Dashboard

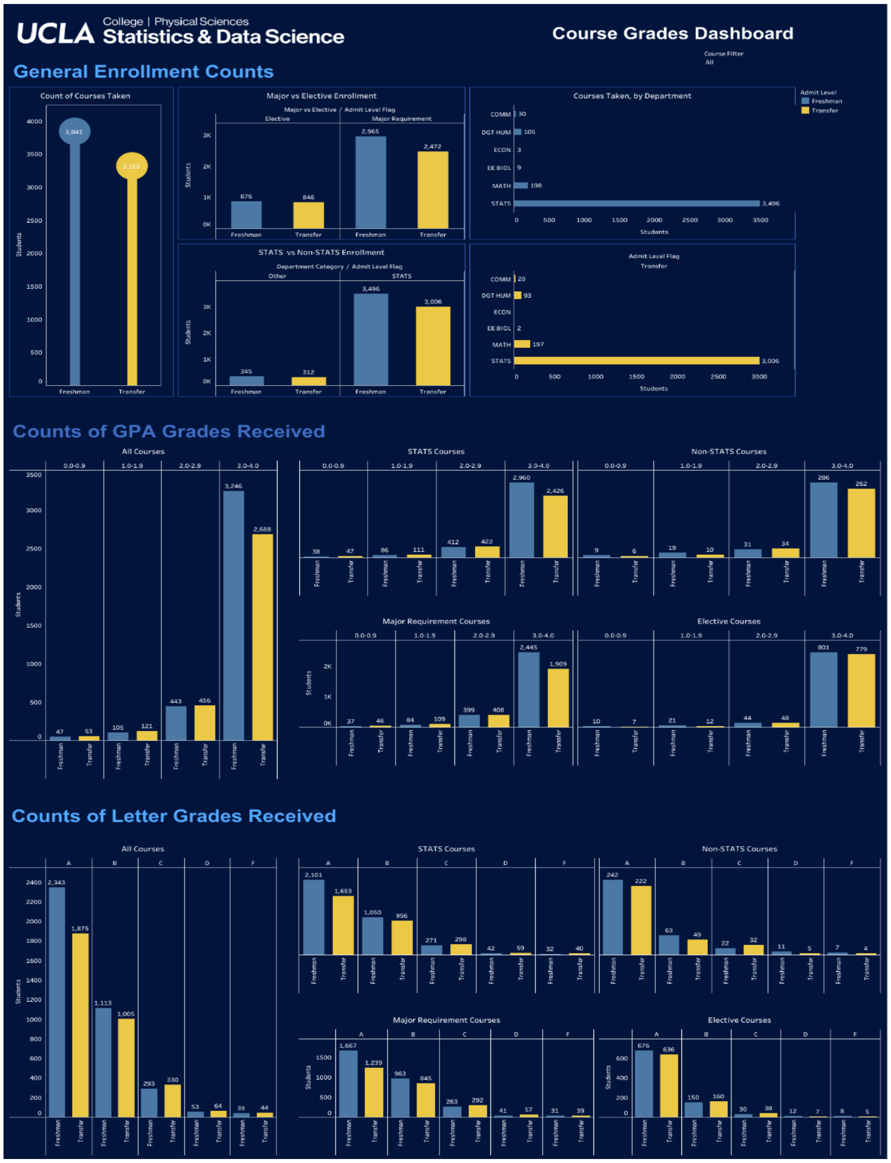

Figure 3. Screenshot of Tableau Dashboard

To further our initial analysis, we created a dashboard utilizing Tableau to provide an interactive visualization of course enrollment patterns and academic performance among Statistics & Data Science students (Figure 3). To prepare the data, we performed cleaning and preprocessing steps, including creating flags for different variables such as admit_level_flag (distinguishing between freshman and transfer admits), gpa_flag (categorizing students into GPA ranges), and grade_flag (grouping letter grades into broader performance categories). We also introduced dimension variables profile_var and var_value to appropriately organize the data for Tableau.

The dashboard aggregates student enrollment counts by course and admission level, allowing users to explore trends in course selection and academic outcomes. The visualization is color-coded, with freshman admits represented in blue and transfer admits in yellow, providing a clear distinction between the two groups. Additionally, users can interact with the dashboard by filtering results based on specific classes, enabling a more focused analysis of student performance in individual courses.

Results

Transfer vs Freshman Students Performance

To compare academic performance between freshman and transfer admits in Statistics courses, we visualized GPA distributions using a KDE plot, allowing for trend analysis independent of class size (Figure 5).

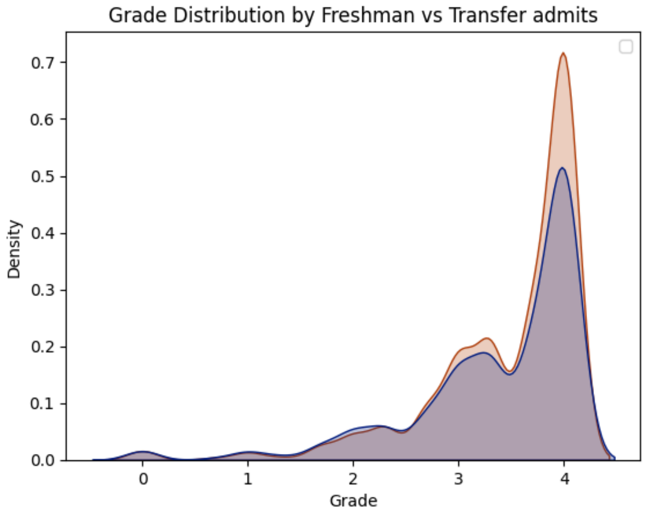

Figure 4. KDE Plot of Grade Distribution for Freshman vs Transfer

The plot shows a left-skewed distribution pattern, with most students having a GPA above a 3.0. Freshman admits demonstrate a stronger presence in the 3.5-4.0 range, shown in orange, suggesting a slightly stronger academic performance in the upper division courses. Additionally, transfer admits seem to have a greater proportion of GPAs below the 3.0 bin.

Since the grade distribution was not normally distributed in the KDE plot, we confirmed this with the Shapiro-Wilk test (Table 2).

Data	Test Statistic	P-value
Freshman	0.743	0.00

Transfer	0.774	0.00

Table 2: Shapiro-Wilk Test For Normality

Since both p-values are less than 0.05, we reject the null hypothesis and confirm that the dataset is not normally distributed.

We then proceed with a one-sided Mann-Whitney U Test with the two following hypotheses:

H_0: There is no significant difference in student performances between transfer and freshman admits

$H_{1:}$ There is a significant difference in student performances between transfer and freshman admits

H_0: There is no significant difference in student performances between transfer and freshman admits

$H_{1:}$ Freshman admits perform better than transfer admits

Alternative Hypothesis	**Test Statistic**	**P-value**
$\mu_{freshman} \neq \mu_{transfer}$	6384555.50	0.00
$\mu_{freshman} > \mu_{transfer}$	6384555.50	0.00

Table 3: One-sided Mann-Whitney U Test

The test results yielded p-values less than 0.05, leading us to reject the null hypothesis and accept the alternative hypothesis, confirming that there is a significant difference in academic performance between the two groups. Transfer students performed worse compared to freshman admits in upper division Statistics courses. These findings highlight potential challenges faced by transfer students as they transition into the Statistics & Data Science major at UCLA.

Course Combinations and GPA Impact

To evaluate how specific course combinations of Statistics and Math courses impact student performance, we construct a pairwise course combination binary matrix. This approach enables us to analyze the impact of specific course interactions on student performance without adding unnecessary feature complexity. Each row represents a student, while each column corresponds to a course. A value of 1 indicates a student's enrollment in that course during the same term and 0 indicates that they did not take the course. To simplify the analysis, we focus on pairwise combinations, which allows us to evaluate how specific course pairings correlate with GPA outcomes.

Using these pairwise combination features, we conduct a feature importance analysis with classification models to quantify how much each course contributes to predicting GPA for a given course pairing.

First, we apply a Random Forest classifier to provide feature importance scores based on how much each course impacts the model's predictions (Figure 5). The top course that plays the most significant role in academic performance is the major-required Stat 100B, Introduction to Mathematical Statistics.

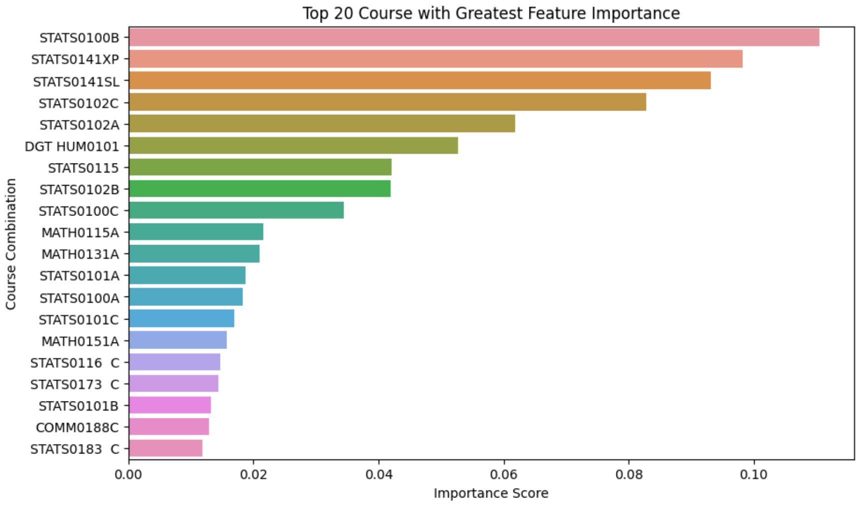

Figure 5. Top 20 Courses By Feature Importance Scores

Additionally, we apply a Logistic Regression model with L1 regularization (Lasso) to assess the individual impact of each course on grades by shrinking less important coefficients to zero, highlighting the most influential courses. Negative coefficients indicate a course is associated with a decrease in the likelihood of earning a higher GPA, while positive coefficients suggest it may increase the probability of achieving a better GPA (Table 4).

Courses Associated with Lower Grades		**Courses Associated with Higher Grades**	
Feature	**Importance**	**Feature**	**Importance**
MATH 182	-1.578393	STATS 116 C	1.110121
STATS 102C	-1.338157	STATS 141SL	0.932680
STATS 100B	-1.248453	MATH 156	0.768524
MATH 135	-1.167014	STATS 115	0.711800
MATH 171	-1.130503	DGT HUM 101	0.512086

ECON 144	-1.085889	STATS 133	0.436742
MATH 170E	-0.939794	MATH 168	0.415585
STATS 143	-0.872327	MATH 131B	0.410051
STATS 102A	-0.829985	STATS 141XP	0.394328
MATH 170B	-0.823253	MATH 151B	0.377294

Table 4. Top Influential Courses from Logistic Regression

By analyzing these coefficients, we identify which courses are most influential in determining student success. We compare the Logistic Regression coefficients with feature importance scores from Random Forest to validate our findings and capture complex relationships between course pairs. This multi-model approach provides a deeper understanding of how different courses and their interactions impact GPA, and uncovers which combinations may be particularly challenging for students and may yield poor grades (Table 5).

Paired Course	**GPA**
[MATH 131, STATS 101C]	2.918
[MATH 115A, STATS 102B]	2.968
[STATS 102B, STATS 161C]	2.985
[STATS 100C, STATS 102C]	2.987
[MATH 115A, STATS 102B]	3.050
[STATS 100B, STATS 101B]	3.055
[STATS 100A, STATS 102A]	3.056
[MATH 170, STATS 101A]	3.108
[STATS 100B, STATS 102A]	3.112
[STATS 102B, STATS 183 C]	3.173

Table 5: Course Combinations Yielding Low GPA

For better visualization, we generated a heat map that highlights the impact of different course combinations on student GPA (Figure 6). This heat map allows us to identify patterns in academic performance by visually distinguishing course pairs that tend to yield higher or lower grades.

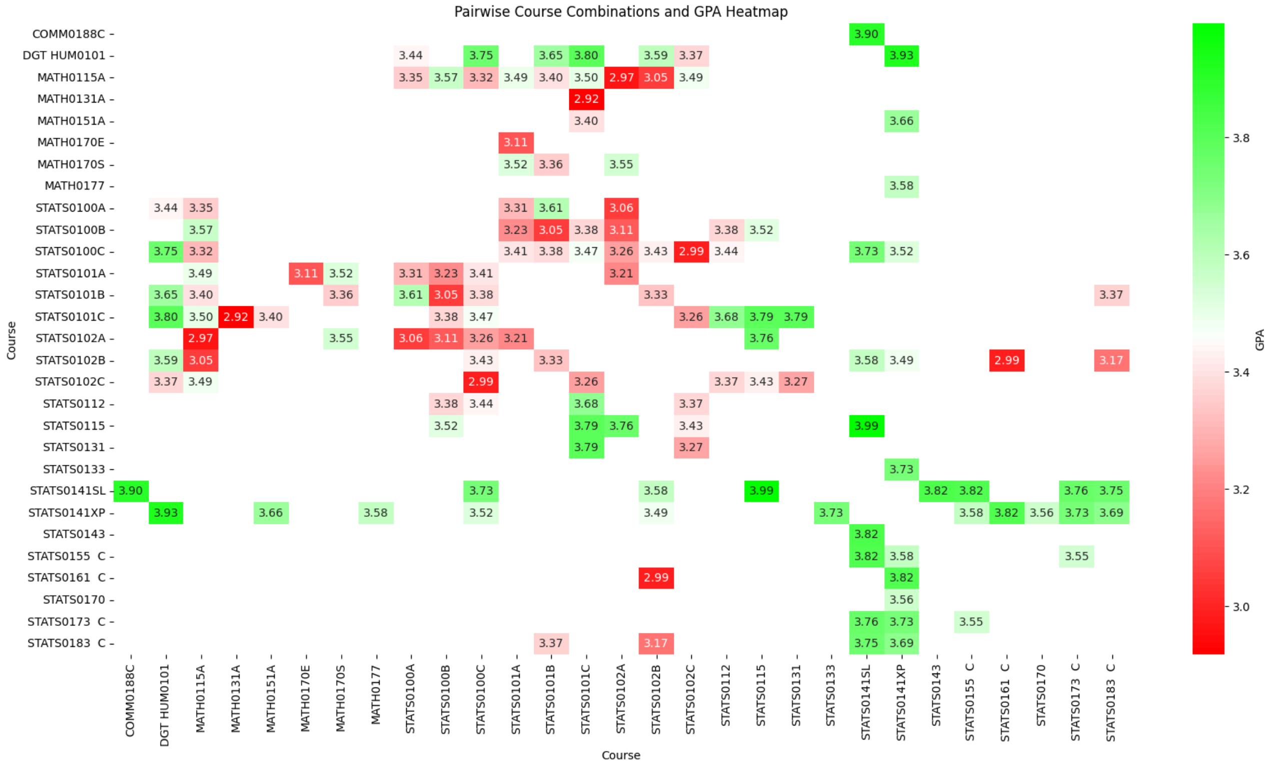

Figure 6: Pairwise Course Combinations and GPA Heatmap

The red areas indicate course combinations associated with lower GPAs, suggesting that these pairs may be especially challenging. For example: [Math 131A, Stats 101C] (GPA 2.918) and [Math 115A, Stats 102B] (GPA: 2.968) are among the lowest-scoring combinations. These findings suggest that students taking these courses together may experience difficulties. On the other hand, green areas represent course pairs with higher GPAs, suggesting that students tend to perform better when enrolled in these combinations. Courses such as Stats 141SL and Stats 115 are associated with stronger academic outcomes, suggesting that they may be aligned with student strengths.

To visualize the effects of the challenging course combinations, we created interaction plots, allowing us to observe how different courses influence grades when taken together. Upon inspection, we notice that challenging course combinations consistently exhibit interactions, as indicated by the absence of parallel lines. This suggests that the impact of one course on grades is dependent on the presence of another.

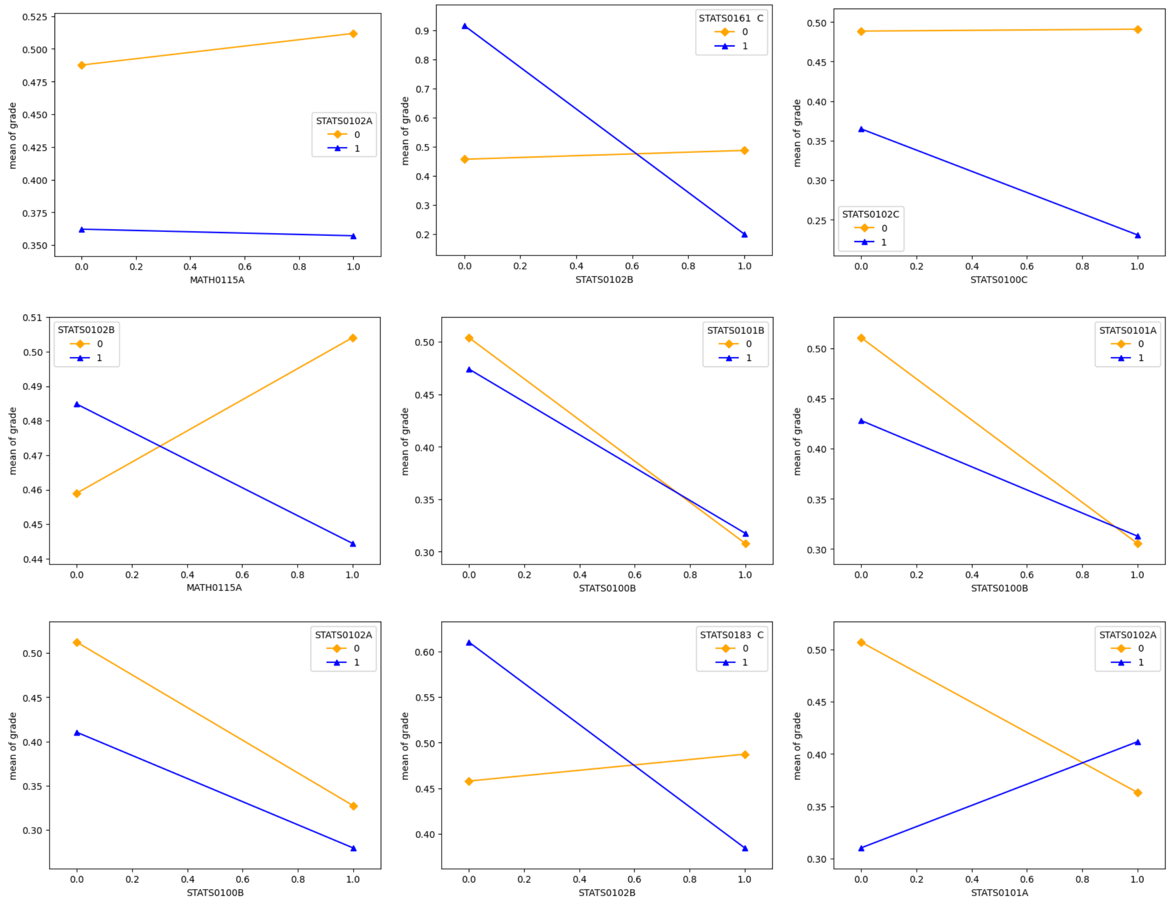

Figure 7: Interaction plots for challenging course combinations

To formally assess the significance of these interactions, we conduct a two-way ANOVA (Analysis of Variance). This statistical test helps determine whether the interaction term between two courses significantly affects grades. We get the following results:

Course	p-value
MATH0115A, STATS0102A	0.517138
MATH0115A, STATS0102B	0.478653
STATS0100B, STATS0101B	1.691757e-01

STATS0100A, STATS0102A	2.439778e-01
STATS0100B, STATS0102A	0.796339
STATS0102A, STATS0101A	**8.050210e-08**
STATS0100C, STATS0102C	**4.406690e-02**
STATS0102B, STATS0183 C	**0.032350**
STATS0100B, STATS0101A	**4.990737e-02**
STATS0102B, STATS0161 C	**0.000074**

Table 6: P-values obtained from Two-Way ANOVA of Challenging Course Combinations

Our two-way ANOVA tests indicate statistically significant interactions between the following course pairs: STATS0102B & STATS0161C, STATS0100C & STATS0102C, STATS0102B & STATS0183C, STATS0101A & STATS0102A, and STATS0100B & STATS0101A. These results suggest that taking these courses together has a distinct impact on performance that goes beyond their individual effects, highlighting the influence of course combinations on student outcomes.

Discussion

Limitations

While this study provides valuable information into academic performance in UCLA's Statistics & Data Science major, several limitations should be considered.

First, the data includes students at different stages in their academic careers, which may influence grade distribution. Students further along in their studies may have developed stronger academic skills, while newer students - particularly transfers - might still be adjusting to the rigor of UCLA courses.

Second, the analysis does not account for differences in professors. Freshman admits - having spent more time at UCLA, may be more familiar with faculty grading styles and could select courses taught by professors who are known for grading less harshly. Transfer students - lacking this institutional knowledge, may not have the same advantage.

Lastly, familiarity with UCLA's academic environment may also play a role. Freshman admits have had more time to adapt to campus resources, quarter systems and support systems, whereas transfer students must navigate these challenges in a shorter period.

Conclusion

The findings of this study reveal significant distinctions in academic performance between freshman and transfer student in the Statistics and Data Science major at UCLA. The analysis of GPA distributions indicates that freshman students generally achieve higher GPAs in upper-division courses compared to transfer students.The Mann-Whitney U test results confirm a statistically significant difference in performance, suggesting that transfer students may face additional challenges when transitioning into upper-division courses.

In conclusion, educational research is essential for improving student learning outcomes. To enhance how students learn, we must first understand their learning styles. Our research focuses on supporting Statistics and Data Science students by identifying key factors that influence their academic success. While the major has core requirements, students have flexibility in choosing the order of their courses and electives. Investigating the best combinations of statistics and math courses can help advisors guide students to have more effective academic pathways, ultimately improving the program's overall efficiency.

References

1. Duggan, M. H., & Pickering, J. W. (2008). Barriers to Transfer Student Academic Success and Retention. Journal of College Student Retention: Research, Theory & Practice, 9(4), 437-459. https://doi.org/10.2190/CS.9.4.c
2. Geiser, S., & Maria Veronica Santelices. (2007). Validity Of High-School Grades In Predicting Student Success Beyond The Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes. UC Berkeley: Center for Studies in Higher Education. Retrieved from https://escholarship.org/uc/item/7306z0zf
3. Schmitz ED. Academic performance of students in lower-division and upper-division environments. [Order No. 28253449]. University of Hawai'i at Manoa; 1984. https://www.proquest.com/dissertations-theses/academic-performance-students-lower-division/docview/2455935763/se-2?accountid=14512
4. Townsend, B. K., McNerny, N., & Arnold, A. (1993). WILL THIS COMMUNITY COLLEGE TRANSFER STUDENT SUCCEED? FACTORS AFFECTING TRANSFER STUDENT PERFORMANCE. Community College Journal of Research and Practice, 17(5), 433–433. doi:10.1080/0361697930170504
5. University of California, Los Angeles. (n.d.-a). First-Year Profile - Fall 2022. Undergraduate Admission. https://admission.ucla.edu/apply/first-year/first-year-profile/2023
6. University of California, Los Angeles. (n.d.-b). First-Year Profile- Fall 2023. Undergraduate Admission. https://admission.ucla.edu/apply/transfer/transfer-profile/2023
7. University of California, Los Angeles. (n.d.-c). Transfer Profile - Fall 2022. Undergraduate Admission. https://admission.ucla.edu/apply/first-year/first-year-profile/2023
8. University of California, Los Angeles. (n.d.-d). Transfer Profile - Fall 2023. Undergraduate Admission. https://admission.ucla.edu/apply/transfer/transfer-profile/2023