California's Unlevel Playing Field

Kylie Scharf[1], Nicole Yee[1], Curtis LeQuang[1], Emma Durler[1]

[1]Department of Statistics, Cal Poly - SLO

May 30, 2024

Abstract

In our report, we investigated whether California's collegiate expenses from 2015 to 2019 are truly greater for baseball than for softball. Under Title IV federal student financial assistance programs, every year colleges across the United States must open their books and reveal the inner workings of their athletic departments. This requirement aims to ensure transparency in athletic funding, participation, and staffing. Since 2003, the Department of Education has collected data from 2,074 schools nationwide through the Equity in Athletics Survey. Focusing on the expenses of specific sports can help us determine if equality in funding is truly ensured. Our research contains a subsetted dataset of 659 California observations from 2015 to 2019. each observation consists of a school in a particular year with a collegiate baseball or softball team. From this data, we created a new variable to represent the difference in funding between baseball and softball for these 659 observations. We created multiple visualizations to better view observed average differences for each year. We then performed a paired t-test to confirm this, finding convincing evidence to conclude that baseball expenses are signifigantly higher than softball expenses from 2015 to 2019 for California colleges like the ones observed. Additional research is needed to explore the reasons behind these disparities and to identify effective strategies for achieving funding equality. Studies could also be expanded to other states and sports to provide a broader understanding of the issue.

1 Introduction

In the sunny state of California, where the crack of the bat and the cheers of fans echo through the air, a tale of two sports unfolds. One is blessed with seemingly unlimited funds and admiration, and the other is left to fend for itself. A stark contrast in funding has long persisted between baseball and softball programs. For decades, college baseball has enjoyed a lion's share of financial support, with state-of-the-art facilities, top-tier coaching staff, and generous scholarship allocations. Meanwhile, softball programs, though equally deserving of investment and recognition, are often overlooked and underfunded.

Dana Drew Shaw (2019) writes of specific examples of this, where many times softball teams are not granted access to locker rooms and are forced to change in the parking lot while baseball teams have locker rooms and offices right next to the fields they play on.

The disparity serves as a barrier to the growth and development of female athletes, perpetuating inequality and hindering their ability to compete at the same level as their male counterparts. Denied access to resources and opportunities, female athletes are at a disadvantage both on and off the field, facing obstacles in their pursuit of academic and athletic success.

In our report, we aimed to answer the question of whether or not the average expenses of collegiate baseball are truly greater than softball in California from 2015 to 2019. The following section introduces the data, subsetting methods, and statistical analyses used to perform our analysis. We then report our findings and discuss the implications of such, Finally, we provide a conclusion and direction for future research.

2 Data and Methods

2.1 Data Collection

The Equity in Athletics Disclosure Act requires all US colleges with a Title IV federal student financial assistance program to report on athletic funding, participation, and staffing for each sport and academic year (Office Of Postsecondary Education 2021). Since 2003, The Department of Education has collected data on collegiate sports budgets at 2,074 schools nationwide in obeyance of the act. The department created the Equity in Athletics Survey to collect such data. Each school that meets the requirements completes this survey annually for each of its athletics teams. We accessed this data through the Equity in Athletics Data Analysis (2003-2023). The obtained data set contains all survey responses from 2015 to 2019. Although the data set contains various types of funding information for all participating schools, we are specifically interested in the expenses for each school's softball and baseball program.

2.2 Statistical Analysis

Prior to analysis, we subsetted the data to include only California schools's softball and baseball programs. We did such using the tidyverse (Wickham et al. 2019) and readxl (Wickham and Bryan 2023) packages. We also used the writexl package (Ooms 2024) to create permanent excel data sets. We then pivoted the data so that each school in each year would represent one observation and contains information about the softball and baseball expenses as such. Our final data set included 659 records of information on the expenses, in dollars, of schools in California from 2015 to 2019 for softball and baseball alike. We also created a new variable to represent the difference in expenses for baseball and

softball (baseball - softball) at each school in each year. We visualized these differences in a table using the kableExtra package (Zhu 2024). We intended to average these differences to determine whether or not these expenses are truly greater for baseball than for softball. To do such, we performed a paired t-test to analyze the average differences over all 5 years and test whether or not they are statistically significant.

We recognize that baseball and softball are not exactly equivalent sports, which will impact some of the differences between the data. We will keep this in mind when completing our data analysis and hope to address as much of this difference as possible.

3 Results

In this report, we wanted to determine if, for the years 2015-2019, California collegiate expenses for baseball were on average greater than softball. California colleges reported their annual individual expenses for baseball and softball teams (n=659). Table 1 breaks down each year's average difference in expenses between the baseball and softball teams of an individual school, standard error of such difference, and number of participating California colleges.

Table 1: The average differences between baseball and softball expenses, the variability in such, and the number of participating California colleges for each year between 2015 and 2019

Year	Difference ($)	Standard Error ($)	Sample Size
2015	126459.8	20152.60	133
2016	147310.8	22632.39	130
2017	150482.4	24847.36	132
2018	168719.2	24457.78	132
2019	156763.0	23088.92	132

Note:

Source: Department of Education

After looking at the table it becomes fairly clear that the expenses for baseball are much higher than the expenses for softball every year. The difference between the sports is over /$100,000 every year (Table 1). Both baseball and softball expenses also appear to be increasing over time from 2015 to 2019 and then decrease in 2019. The variability, in terms of standard error, of the differences appears to remain relatively constant but slighlty increases from 2015 to 2018 and then similarly decreases in 2019. The sample sizes also remain relatively constant, representing 130 to 133 schools per year (Table 1).

After performing a paired t-test, we found evidence that baseball expenses were discernibly greater than softball ($t = 14.542$; $df = 658$, $p < 0.0001$). Baseball ex-

penses averaged about $149,919 more than softball expenses (95% CI: 129676.6, 170162.3). More specifically, we are 95% confident that at the collegiate level, baseball expenses were, on average, between $129,676 and $170,162 greater than softball expenses.

More prominently displayed in Figure 1, we can see that the average differences in expenses for a particular year are well above 0. The highlighted section in the figure represents a roughly 95% confidence interval for the true difference in expenses, where the width is two standard errors of the average differences, or $20,618. The entirety of the interval is clearly above 0 in the visualization.

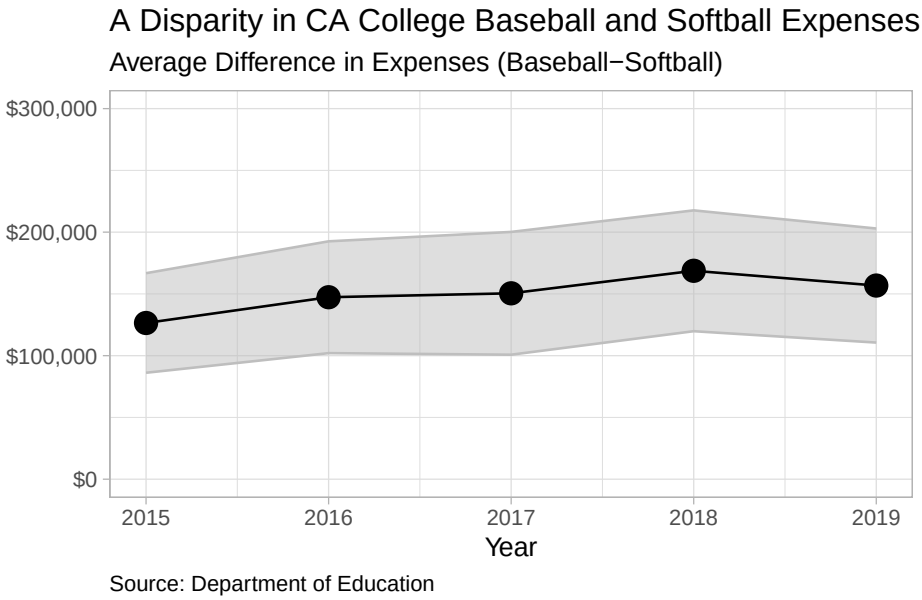

Figure 1: Above entails an analysis of the average differences in expenses from 2015 to 2019 for collegiate baseball and softball in California. The average difference in expenses for a particular year for all California teams is plotted as a black circle. The grey highlighted region represents adding or subtracting two standard errors of the average difference from the plotted average differences. This shaded region therefore represents a roughly 95% confidence interval for the true average difference for all collegiate baseball and softball teams in California between 2015 to 2019 that meet the conditions required for data collection. The average difference in expenses are consistantly greater than 0 along with the interval for said average differences, meaning the average expenses for baseball are consistantly signifigantly larger than for softball.

4 Discussion

Our analysis consists of a paired t-test performed in R that compares the difference in expenses between collegiate baseball and softball in California from 2015 to 2019. Given the t-value of 14.542 and p-value of < 0.0001 (where degrees of freedom = 658), we have convincing evidence to reject the null hypothesis. In context, we have statistically significant evidence to support the alternative hypothesis that the true mean difference in expenses between collegiate baseball and softball is not equal to 0, meaning baseball expenses are significantly greater than softball expenses. However, one limitation of this test is that we cannot assume causation from these results as our data came from an observational experiment. This also means we can only generalize these results to colleges/universities similar to those included in the study. Looking at our graph, we can see that the differences each year are greatly above 0, further supporting our alternative hypothesis. Regarding variability (represented by the shaded region of the graph), we see that each year has roughly the same amount of variability in the difference of the expenses. Lastly, the table shows how much higher average baseball expenses are each year compared to softball, further demonstrating how the difference between the two is large each year and therefore greater than 0.

5 Conclusion

Through our investigation of the difference in expenses between collegiate baseball and softball, we found that baseball consistently has higher average expenses in California over the years. This allowed us to conclude that the true mean difference between the two is not equal to 0. We created multiple visualizations, such as a line graph and a table, and performed a paired t-test using the expenses of softball and baseball teams in California from 2015-2019 in order to come to this conclusion. While we have extremely strong evidence of this statistically, the visual trends in our results provide further evidence. We were also able to see a similar trend in their difference over time as the magnitude of average baseball expenses minus average softball expenses was very consistent throughout every year. For further research, we wonder if this pattern of the difference of average expenses between the two exists through 2024, but since our data only includes data up to 2019, we can only assume. Additional research is needed to explore the reasons behind these disparities and to identify effective strategies for achieving funding equality. Studies could also be expanded to other states and sports to provide a broader understanding of the issue. More research could also be conducted on the decrease in funding obverved in 2019 for both baseball and softball.

References

U.S Department of Education. (2003-2023). *Equity in Athletics Data Analysis*. https://ope.ed.gov/athletics/#/datafile/list

Office Of Postsecondary Education. 2021. "TREND: Office of Postsecondary Education. College Athletics: Total Participation | State: California | Educational Institution: Loyola Marymount University | Collegiate Sport: Rowing | Variable: Total # Participants Women, 2005 - 2018. Data Planet™ Statistical Datasets: A SAGE Publishing Resource Dataset-ID: 069-001-001." https://doi.org/10.6068/DP176F7A216C817.

Ooms, Jeroen. 2024. "Writexl: Export Data Frames to Excel 'Xlsx' Format." https://CRAN.R-project.org/package=writexl.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the {Tidyverse}" 4: 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Jennifer Bryan. 2023. "Readxl: Read Excel Files." https://CRAN.R-project.org/package=readxl.

Zhu, Hao. 2024. "kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax." https://CRAN.R-project.org/package=kableExtra.