

Data Description

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. Since 2003, The Department of Education has collected data on collegiate sports budgets at 2,074 schools nationwide in obedience 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, a dataset containing all survey responses. A single observation in the dataset represents a specific collegiate sport in one particular year. As softball and baseball are considered two different sports, they each have their own row. The dataset includes the sport, school, city, state, number of participating men, number of participating women, revenue from men, revenue from women, expenses for men, and expenses for women. We decided to subset this data to include only data from California schools's softball and baseball programs to make the analysis more manageable. We are additionally only using the period 2015-2019 for the same reason. We chose California because we attend college in California and chose baseball and softball because we are all very interested in analytics related to such sports. 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.

References

- <https://usafacts.org/articles/coronavirus-college-football-profit-sec-acc-pac-12-big-ten-millions-fall-2020/>
- <https://ope.ed.gov/athletics/#/>
- <https://ope.ed.gov/athletics/#/datafile/list>
- <https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-29/readme.md> (data set)

Citations

Coronavirus and college football: what might this mean for colleges' bottom lines? (2020,

September 23). USAFacts; USAFacts.

<https://usafacts.org/articles/coronavirus-college-football-profit-sec-acc-pac-12-big-ten-millions-fall-2020/>

Equity in Athletics. (n.d.-a). Ope.ed.gov. <https://ope.ed.gov/athletics/#/>

Equity in Athletics. (n.d.-b). Ope.ed.gov. <https://ope.ed.gov/athletics/#/datafile/list>

Tidy Tuesday Data Sets. (2022, March 29). GitHub.

[https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-29/readme
_md](https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-29/readme.md)

Link to Data:

<https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-29/readme.md>

Research Questions

1. How does the revenue generated by collegiate baseball and softball differ between California conferences?
 - a. Tentative analysis plan: (1) subset to only softball and baseball and only California (s) create a conferences variable as one does not exist yet (3) perform an ANOVA for the different levels of the conferences and their revenues (hmms) we may need to pivot data to create the conference variable
2. Is there a difference in the relationship between expenses and revenue for collegiate softball and baseball in California?
 - a. Tentative analysis plan: (1) subset to only softball and baseball and only California (2) two-sample t-test for slopes (hmms) we may need to use a different analysis than a two-sample t-test depending on the way we look at this question
3. Is there a significant difference between the revenue generated by California collegiate baseball and softball?
 - a. Tentative analysis plan: (1) subset to only softball and baseball and only California (2) two sample t-test using the revenue variable for men vs. for women (hmms) we may want to average over years or conferences

Group Contract:

We will communicate via text and in person because we share a lot of common classes. We will expect updates daily (mainly because we see each other often) depending on where we are in the project and what deadlines are coming up throughout the quarter. We will not necessarily have a "leader" as we expect to rotate and share the work evenly. As for a timeline of completion, we will try to keep up accordingly with each mini-project deadline by communicating frequently and meeting in person when required. We plan to use GitHub to work collaboratively on the project. We do not plan to assign different parts of the project to specific group members, rather we plan to work together as one to be more on the same page throughout our busy quarters. If someone is sick or super busy, we expect that person just to reach out, and whoever is willing/free can cover accordingly. If we have any group conflicts (although we do not expect this to happen), we will come together as a group to solve the problem, but if it continues/worsens, we may inform Professor Robinson and suggest point deductions.