

Honors Applied Statistics Final Project Proposal

Troy Edwards

2023-04-14

1 Research question

The question that I seek to answer with my research is: “What factors influence an FRC team’s success?” I chose this question because I myself am on an FRC team and I am interested to see what factors influence success to try to improve my own team’s success.

2 Related literature

There is not much work that has already been done on this topic. I managed to find 2 papers: “An Analysis of the Success of FRC Robotics Teams” by Max Tepermeister and “An Overview and Analysis of Statistics used to Rate FIRST Robotics Teams” by William Gardner. The first one is not a paper that I would consider “scholarly” because it is a high-schooler’s final project for AP Statistics, but the second one appears formal and well-organized. However, Gardner’s paper is not fully on-topic, as it analyzes metrics used to rate FRC teams. The information presented is still relevant, just not directly applicable. Tepermeister’s objective was similar to mine. Tepermeister mainly analyzed the effects of the age, size, and budget of a team on its success. Tepermeister found that none of these factors have a significant effect on success.

3 Gaps in current research

Tepermeister’s research only analyzed a few factors, and was more of a general analysis and did not address performance in specific games, where robot design can have a huge influence on performance. I seek to build on Tepermeister’s research with similar analysis, along with new research analyzing teams’ capabilities in robot design in the recent FRC games RAPID REACT (2022) and CHARGED UP (2023). Gardner’s research only analyzed the effectiveness of various techniques and metrics used to rate FRC teams and measure their performance. I can use Gardner’s research to decide what methods to use to measure success.

4 Data sources

I plan to use Evan Kuykendall's 2023 scouting data (not publicly available to my knowledge) and Jake Benjamin's 2022 scouting data (also not publicly available to my knowledge). I also plan to possibly use data from The Blue Alliance, Statbotics, and the FRC Events API.

5 Research methods

I plan to take a random sample of 30 teams from the Pacific Northwest district for each year that I analyze. I will use teams from the Pacific Northwest for a more in-depth analysis because the data that I have access to is much more in-depth and easy to use for this region.

6 Predictive models and inference tests

I plan to do use the same process to analyze both RAPID REACT and CHARGED UP for the Pacific Northwest. This process includes 3 main steps:

1. Run ANOVA tests and/or t -tests to determine whether various robot features have a significant effect on a robot's success.
2. Create a multiple linear regression model to predict a team's win rate for the season.
3. Create a multiple logistic regression model to predict the probability of a specific team winning in any given match.

7 Expected challenges

I am expecting that there will be a couple of teams in the Pacific Northwest that will not have been included in the scouting data because they were not at the events where the data was gathered. This has the potential to make it more difficult to analyze specific teams that are not at the events where the data was gathered. To avoid this, I will only select my sample from teams that were at events where the data was gathered. This should not introduce too much bias because the main factor that influences which events teams go to is geographic location. Another challenge is that many teams may not disclose information such as team size and age. This will likely influence whether or not I analyze these factors. If I do, unavailable data will be treated as NA and ignored.