Analysis of the Various Factors Influencing the Success of FRC Teams

Troy Edwards

2023–06–16

Contents

Abstract

TODO

1 Introduction

FIRST Robotics Competition is a robotics competition in which teams of high schoolers build robots to compete in a game that is different every year.. FIRST is the organization that runs these competitions, and the acronym FIRST stands for "For the Inspiration and Recognition of Science and Technology." Each year, FIRST reveals a new game in which alliances of 3 industrial-size robots compete against each other on a field to complete various tasks. Teams have limited time after "Kickoff" (the game reveal) to design, build, wire, and program their robot to perform the tasks required for the game. For example, in this year.'s game, CHARGED UP, robots must travel across the field to the Substation Area to pick up one game piece (a cube or a cone) at a time and then travel back to their alliance's Grid and deposit the pieces onto Nodes.

There are many factors that could have an effect on a team's success. These include general things about the team, such as the team's age, its budget, and its size, and also things about their robot in a specific season, such as drivetrain type and scoring capability. The goal of this research is to determine which of these factors have the greatest effect on a team's success.

2 Literature Review

There is not much work that has been done on this topic. The most relevant paper that I could find was "An analysis of the Success of FRC Robotics Teams" by Max Tepermeister. Tepermeister found that team age had a small amount of correlation with success, while team budget and team size had no correlation with success. However, Tepermeister used OPR (offensive power rating) rather than win rate to measure a team's success, which means that his results could be different from mine.

3 Method

3.1 Analysis of individual variables

I plan to analyze whether several variables have a significant influence on a team's success. Only qualification matches will be analyzed because the alliances are random and the data only covers scouting matches. **TODO: ADD MORE HERE (HOW IS SUCCESS MEASURED?)**

3.1.1 Drivetrain

A robot's drivetrain can have a huge influence on how successful it is. FRC games are typically quite fast-paced, so a better drivetrain can make a huge difference. For most games, the playing field is quite flat, with few obstacles to overcome. This is true of both RAPID REACT and CHARGED UP. In games with few obstacles, teams will typically use some sort of "normal" drivetrain (i.e. one without huge wheels or anything like that). The three main types of drivetrain typically found in FRC are:

1. Differential
2. Mecanum
3. Swerve

A differential drive has normal wheels on the left and right sides of the robot. A Mecanum drive uses Mecanum wheels, which are special wheels with diagonal rollers mounted around the circumference. A robot with a mecanum drivetrain can drive forwards and backwards, similar to a differential drivetrain, but it can also drive sideways by driving specific wheels in specific directions. A swerve drive is another type of drivetrain that can drive in any direction, but unlike the Mecanum drivetrain, the swerve drive achieves this by using two motors for each wheel: one to drive the wheel, and another to steer the wheel. This allows each wheel to be steered independently of the others, which allows for complex maneuvers.

These can be grouped into two categories, namely holonomic (Mecanum and swerve) and non-holonomic (differential). The difference between these two categories is that while a non-holonomic drivetrain constrains a robot to only move forwards and backwards such that it must turn to change its direction of motion, a holonomic drivetrain is free of these constraints.

A differential drive is usually the cheapest, since it requires no special parts. The next cheapest of these three is the Mecanum drive because you only need the 4 wheels and a motor for each. However, one major drawback of a Mecanum drivetrain is that the wheels are more prone to slipping. This means that a robot with a Mecanum drivetrain has a lower maximum acceleration before the wheels start to slip. It also means that in FRC, a robot with a Mecanum drivetrain is very easy to play defense against because a robot with a differential drive or a swerve drive can easily shove it it around. Finally, the most expensive, but also the most effective is the swerve drive. The swerve drive combines the lack of slip (and by extension, pushing power) of a differential drivetrain with the flexibility of a holonomic drivetrain. The main drawback of swerve drive is that it can at times be prohibitively expensive, with a single swerve module typically costing upwards of $300 not even including the motors. This is because they are very complex, with lots of gears and custom machined parts.

To analyze the effect of drivetrain on a team's success, I will perform a two sample difference of means t-test to compare the win rate of teams with holonomic drivetrains to teams without holonomic drivetrains.

3.1.2 School budget

A school's budget can have a huge influence on a team's budget, which can in turn influence the team's success by allowing them to use more complex tools, purchase higher-end parts, and go to more events.

To analyze the effect of budget on a robot's success, I will create a model using school expenditures per pupil to predict win rate and perform a t-test for slope to see if the relationship is significant.

3.1.3 Team age

An older team will likely have more advanced techniques and manufacturing processes at its disposal, while a younger team will typically not have these things. Team number is a good proxy to team age because the team numbers are assigned sequentially.

To analyze the effect of a team's age on its success, I will create a model using team age to predict win rate and perform a t-test for slope to see if the relationship is significant.

3.2 Predictive models

After analyzing individual variables, I will create a multiple linear regression model using a combination of the above variables to predict a team's overall win rate

3.3 Data collection and sources

My data comes from three main sources:

1. Evan Kuykendall's 2023 scouting data from Glacier Peak, Auburn, and the Pacific Northwest District Championship
2. Jake Benjamin's 2022 scouting data from Auburn and the Pacific Northwest District Championship
3. The Blue Alliance

The scouting datasets contain information about each robot in every match that was scouted. An example of something that would be included in the data for 2022 would be number of pieces scored in the UPPER HUB while an example of something that would be included in the data for 2023 would be number of pieces scored in HYBRID NODES. The Blue Alliance only has more general data, such as match results (win/loss). However, it also contains information about who was on what alliance for a specific match. This will be useful for gathering data about alliance members to use in the logistic regression model. I also have data from the Department of Education for the inference test with school expenditures.

4 Findings

4.1 Tests

4.1.1 Win rate vs. drivetrain type

$$\text{Let } \mu_h = \text{True mean win rate of teams whose robots have holonomic drivetrains}$$

$$\text{Let } \mu_{nh} = \text{True mean win rate of teams whose robots have non-holonomic drivetrains}$$

$$H_0 : \mu_h = \mu_{nh}$$

$$H_a : \mu_h \neq \mu_{nh}$$

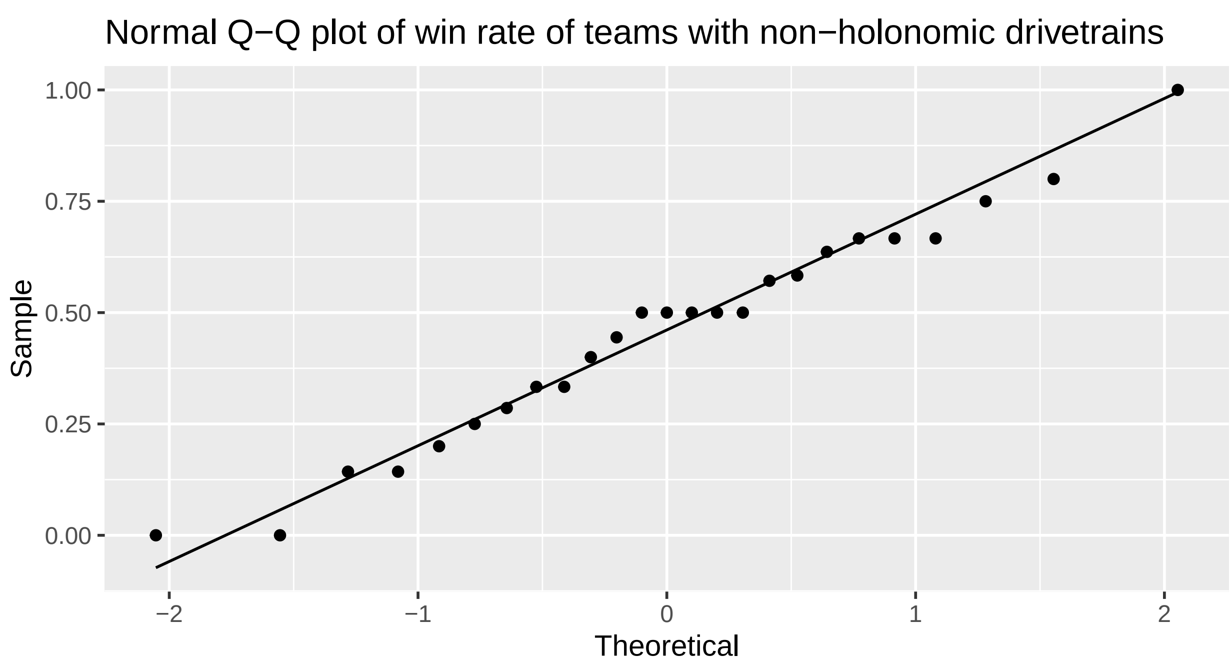
$$\alpha = 0.05$$

Two-sample difference in means t-test

Random: teams were selected using a simple random sample
Normal:

- $n_h = 34 \Rightarrow$ Central limit theorem states that the distribution of sample means is approximately normal

- $n_{nh} = 25 \Rightarrow$ CLT fails, need normal probability plot




```
##
##  Welch Two Sample t-test
##
## data:  win_rate by dt_type
## t = 1.4802, df = 53.038, p-value = 0.1447
```

```
## alternative hypothesis: true difference in means between group h and group nh is not
## 95 percent confidence interval:
##  -0.0354328  0.2350284
## sample estimates:
##  mean in group h mean in group nh
##        0.5547444        0.4549466
```

Normal probability plot is approximately linear $\Rightarrow$ win rate of teams with non-holonomic drivetrains is approximately normally distributed

Independent:

- $n_h = 34 \not< 0.1N_h$, proceeding with caution
- $n_{nh} = 25 \not< 0.1N_{nh}$, proceeding with caution

$$t = 1.4802$$

$$p = 0.1447 \text{ on } 53.038 \text{ df}$$

Since our p-value (0.1447) is greater than α (0.05), we fail to reject our null hypothesis. Therefore, we do not have enough evidence to support the alternative hypothesis.

4.1.2 Win rate vs. school expenditure per pupil

Let $\beta =$ True mean slope of the relationship between win rate and school expenditure per pupil

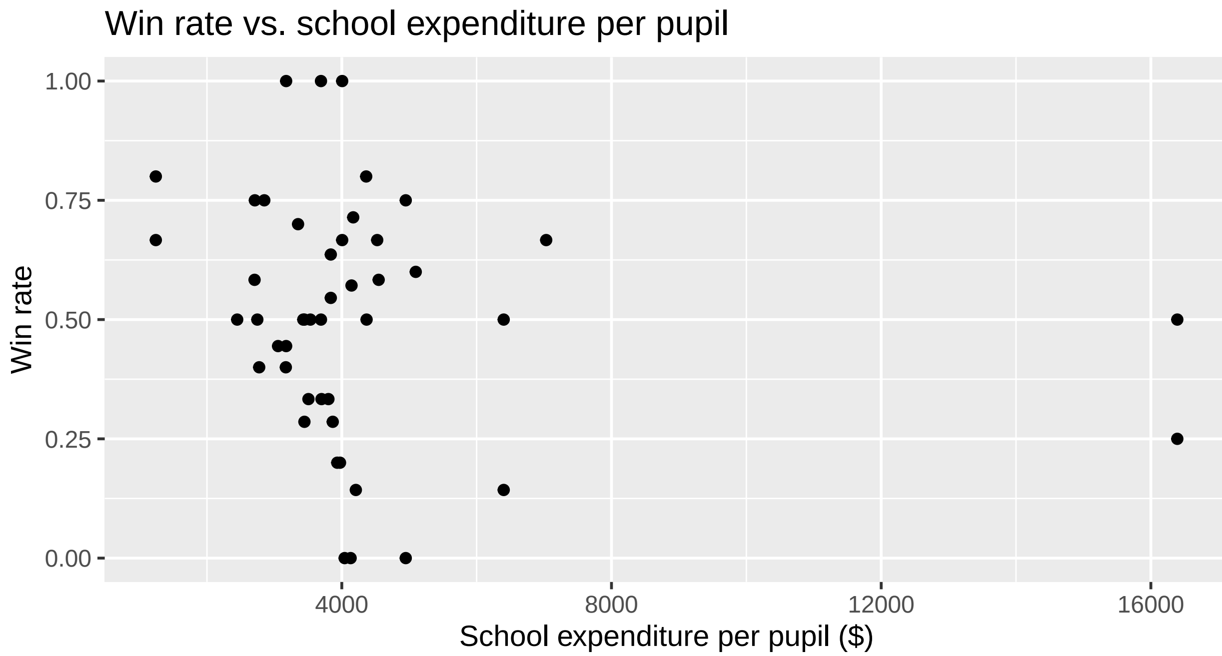
$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

$$\alpha = 0.05$$

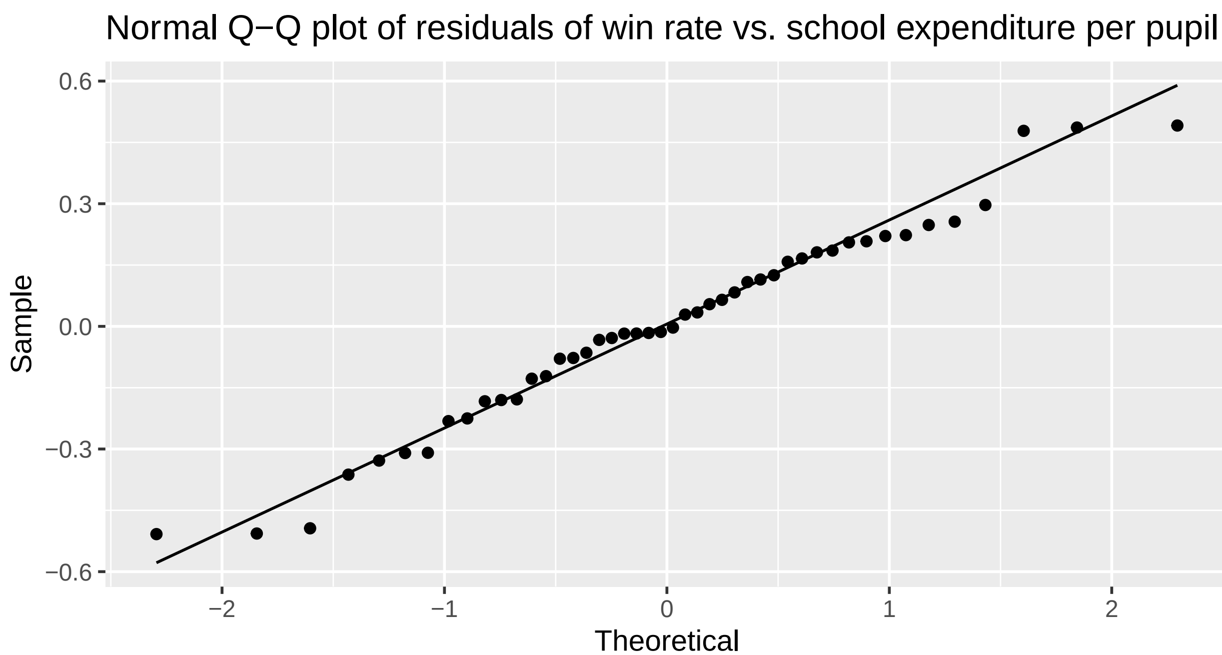
$$t\text{-test for slope}$$

Linear:



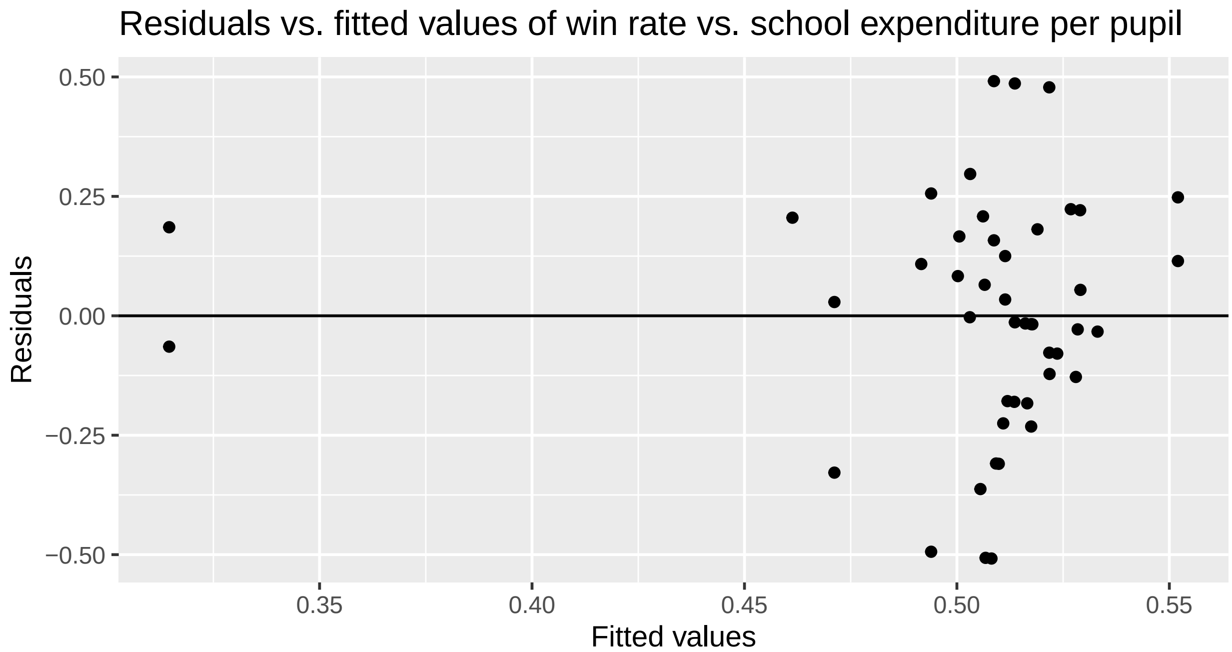
Scatterplot of the data does not appear linear ⇒ proceeding with caution

Normal:



Q-Q plot is approximately linear ⇒ residuals are approximately normally distributed

Equal Variance:



Residuals appear to be randomly distributed around 0 $\Rightarrow$ variance is approximately equal

```
##
## Call:
## lm(formula = win_rate ~ eps, data = team_data)
##
## Coefficients:
## (Intercept)          eps
##   5.715e-01   -1.567e-05
```

$$t = -1.189$$

$$p = 0.241 \text{ on } 44 \text{ df}$$

Since our p-value (0.241) is greater than α (0.05), we fail to reject our null hypothesis. Therefore, we do not have enough evidence to support the alternative hypothesis.

4.1.3 Win rate vs. team number

$$\text{Let } \beta = \text{True mean slope of the relationship between win rate and team number}$$

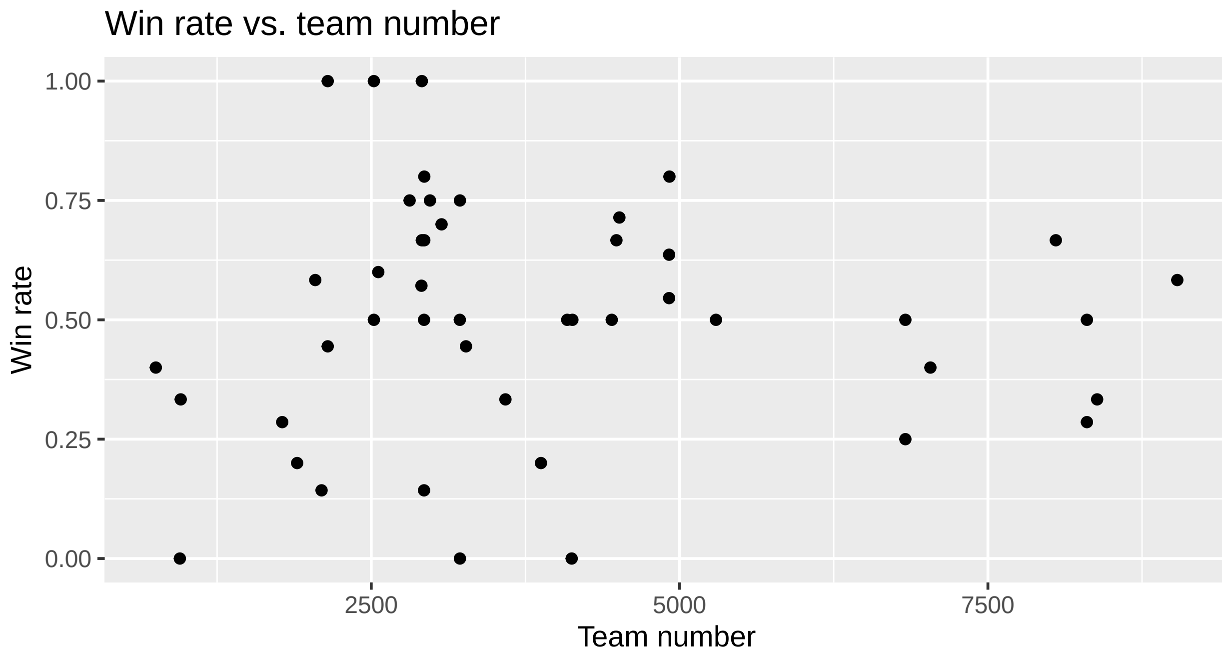
$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

$$\alpha = 0.05$$

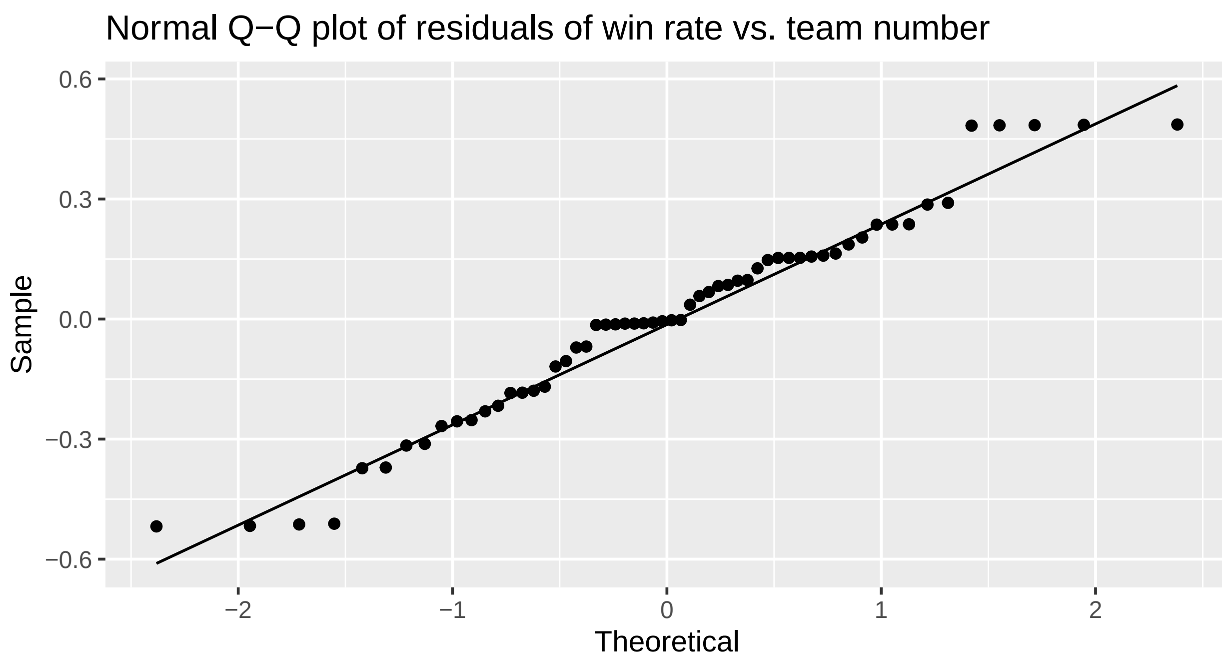
$$t\text{-test for slope}$$

Linear:



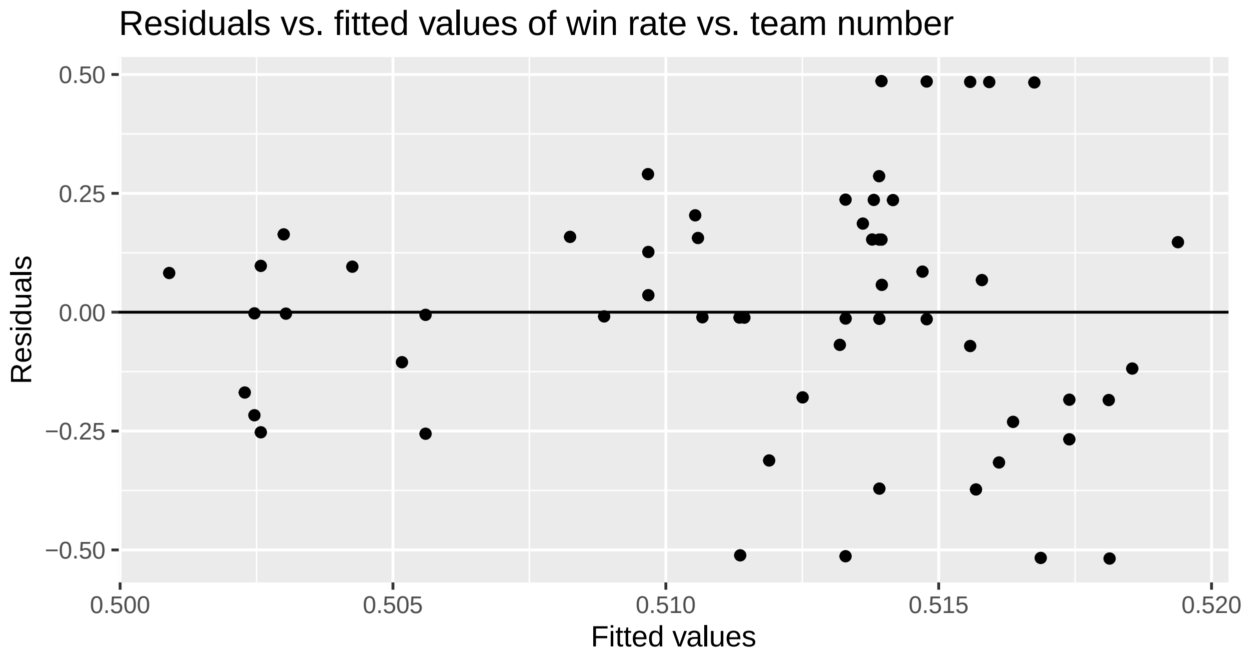
Scatterplot of the data does not appear linear ⇒ proceeding with caution

Normal:



Q-Q plot is approximately linear ⇒ residuals are approximately normally distributed

Equal Variance:



Residuals appear to be randomly distributed around 0 $\Rightarrow$ variance is approximately equal

```
##
## Call:
## lm(formula = win_rate ~ team, data = team_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51813 -0.18284 -0.00432  0.15528  0.48605
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.202e-01  6.760e-02   7.694 2.46e-10 ***
## team        -2.131e-06  1.474e-05  -0.145    0.886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2608 on 56 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.0003729,  Adjusted R-squared:  -0.01748
## F-statistic: 0.02089 on 1 and 56 DF,  p-value: 0.8856
```

$$t = -0.145$$

$$p = 0.886 \text{ on } 56 \text{ df}$$

Since our p-value (0.886) is greater than α (0.05), we fail to reject our null hypothesis.

Therefore, we do not have enough evidence to support the alternative hypothesis.

4.2 Predictive Model

Let

$$\hat{y} = \text{Predicted win rate}$$

$$x_1 = 1 \text{ if team has a non-holonomic drivetrain, 0 if team has a holonomic drivetrain}$$

$$x_2 = 1 - x_1$$

$$x_3 = \text{School expenditure per pupil}$$

$$x_4 = \text{Team number}$$

Model	r^2	r^2(adj)	F
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_4$	0.1256	0.06319	2.012
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3$	0.08679	0.04432	2.043
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_4$	0.03889	0.003939	1.113
$\hat{y} = \beta_0 + \beta_1 x_3 + \beta_2 x_4$	0.03499	-0.009898	0.7795
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_4 + \beta_4 x_1 x_3 + \beta_5 x_1 x_4 + \beta_6 x_3 x_4 + \beta_7 x_1 x_3 x_4$	0.2099	0.06437	1.442
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3 + \beta_4 x_2 x_4 + \beta_5 x_1 x_4 + \beta_6 x_2 x_3 x_4 + \beta_7 x_1 x_3 x_4$	0.2099	0.06437	1.442
$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_2 x_4 + \beta_4 x_1 x_4 + \beta_5 x_2 x_3 x_4 + \beta_6 x_1 x_3 x_4$	0.1991	0.0759	1.616

4.2.1 The model

```
##
## Call:
## lm(formula = win_rate ~ dt_type + eps + dt_type:eps + dt_type:team +
##     dt_type:eps:team, data = team_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61149 -0.13080 -0.00288  0.12512  0.46556
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.709e-01  3.545e-01   1.893   0.0661 .
## dt_typenh           4.599e-01  5.962e-01   0.771   0.4453
## eps                -7.941e-05  9.244e-05  -0.859   0.3957
## dt_typenh:eps      -1.023e-04  1.420e-04  -0.721   0.4756
## dt_typeh:team       3.722e-05  7.292e-05   0.510   0.6127
## dt_typenh:team     -1.234e-04  8.648e-05  -1.427   0.1617
```

```
## dt_typeh:eps:team   6.757e-09  1.417e-08   0.477   0.6362
## dt_typenh:eps:team  3.287e-08  1.934e-08   1.700   0.0973 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2418 on 38 degrees of freedom
##   (14 observations deleted due to missingness)
## Multiple R-squared:  0.2099, Adjusted R-squared:  0.06437
## F-statistic: 1.442 on 7 and 38 DF,  p-value: 0.2175
```

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_3 + \beta_3 x_1 x_3 + \beta_4 x_2 x_4 + \beta_5 x_1 x_4 + \beta_6 x_2 x_3 x_4 + \beta_7 x_1 x_3 x_4$$

where

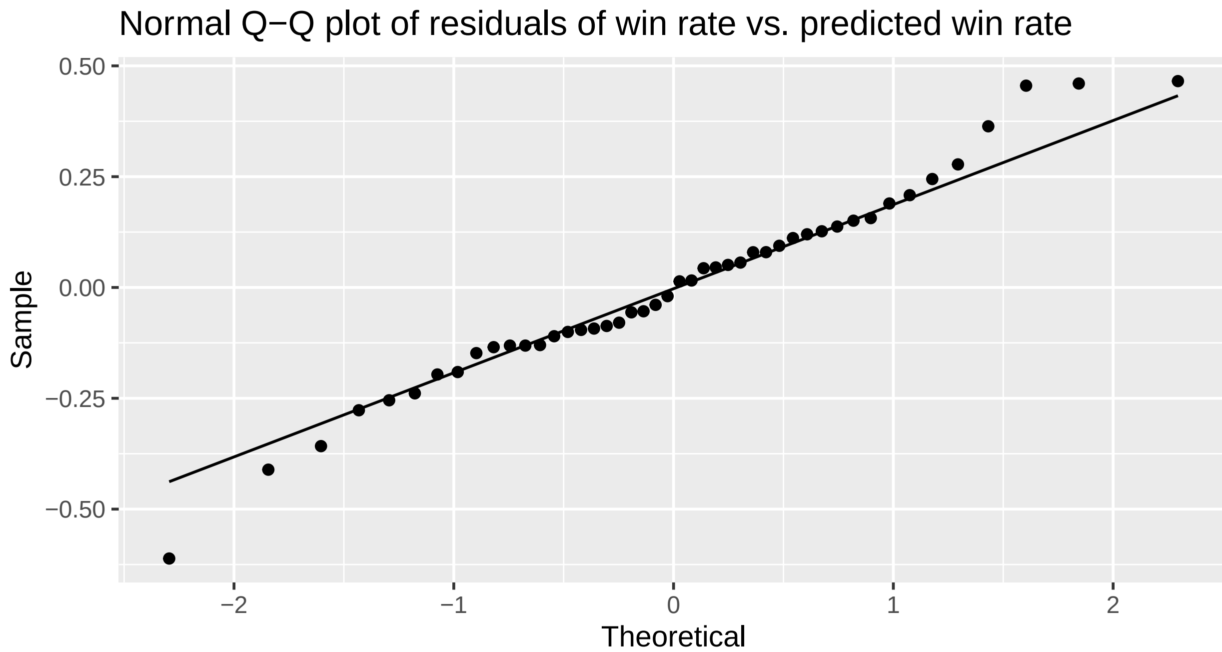
$$\beta_0 = 0.6709$$
$$\beta_1 = 0.4599$$
$$\beta_2 = -0.00007941$$
$$\beta_3 = -0.0001023$$
$$\beta_4 = 0.00003722$$
$$\beta_5 = -0.0001234$$
$$\beta_6 = 0.000000006757$$
$$\beta_7 = 0.00000003287$$

4.2.2 F-test for model utility

$$H_0 : \sum_{i=1}^{7} \beta_i^2 = 0$$
$$H_a : \sum_{i=1}^{7} \beta_i^2 > 0$$
$$\alpha = 0.05$$

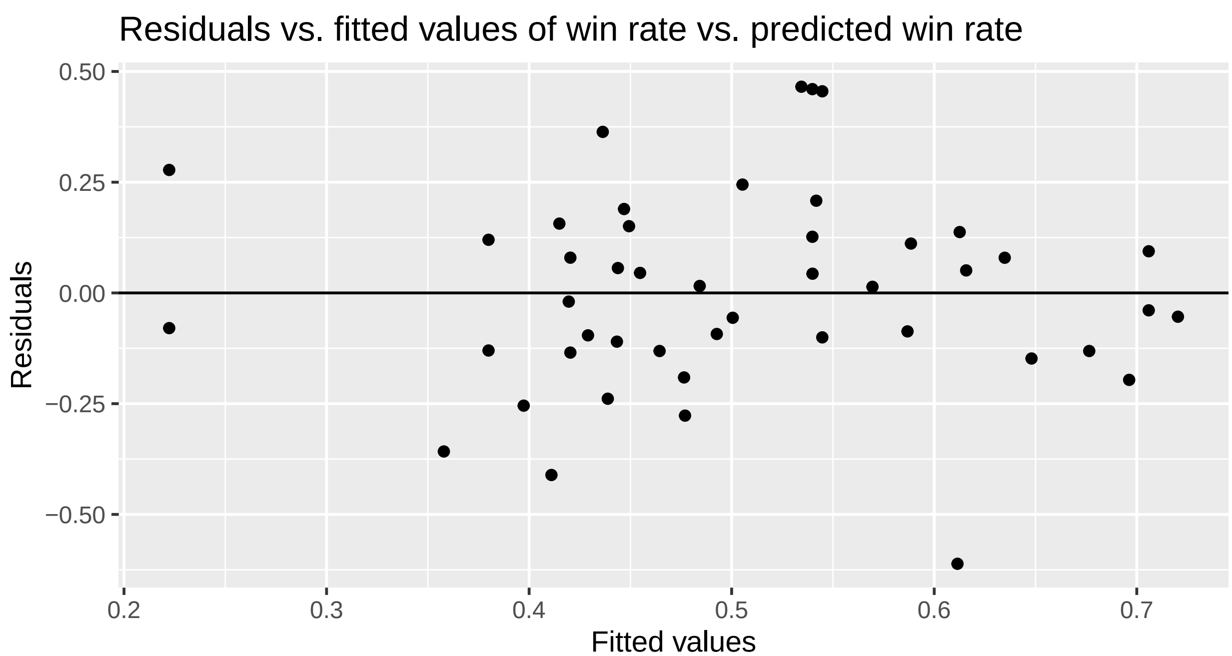
F-test for model utility

Normal:



Q-Q plot is approximately linear $\Rightarrow$ residuals are approximately normally distributed

Equal Variance:



Residuals appear to be randomly distributed around 0 $\Rightarrow$ variance is approximately equal

$$F = 1.442$$

$$p = 0.2175 \text{ on } 7 \text{ and } 38 \text{ df}$$

Since our p-value (0.2175) is greater than α (0.05), we fail to reject our null hypothesis. Therefore, we do not have enough evidence to support the alternative hypothesis.

5 Analysis

5.1 Drivetrain

I thought that drivetrain would be a good predictor of win rate because it is a good indicator of a team's experience and skill. I thought that teams with holonomic drivetrains would have higher win rates than teams with non-holonomic drivetrains because holonomic drivetrains are more complex and require more experience to build and program. However, the p-value of the t-test for drivetrain was 0.886, which is much greater than α. This indicates that drivetrain is not a good predictor of win rate.

5.2 School Expenditure per Pupil

School expenditure per pupil could have been a good predictor of win rate because it is a good indicator of a team's budget and resources. I thought that teams with higher school expenditure per pupil would have higher win rates because they would have more money to spend on their robots. However, the p-value of the slope t-test for school expenditure per student was 0.241, which is greater than α. This indicates that school expenditure per pupil is not a good predictor of win rate.

5.3 Team age

Team number is a good proxy for team age because team numbers are assigned sequentially. I thought that team age would be a good predictor of win rate because it is a good indicator of a team's experience and skill. I thought that older teams would have higher win rates than younger teams because older teams would have more experience. However, the p-value of the slope t-test for team age was 0.886, which is much greater than α. This indicates that team age is not a good predictor of win rate.

5.4 Predictive Model

The predictive model combined the three predictors listed above into a single model. I thought that this model would be a good predictor of win rate because it combines the three predictors listed above. However, the p-value of the F-test for model utility was 0.2175, which is much greater than α. This indicates that the predictive model is not effective at predicting win rate.

6 Conclusion

In conclusion, I was not able to find any good predictors of win rate. This is likely because win rate is a poor metric for measuring a team's performance. Win rate is a poor metric because it is not a good indicator of a team's skill. For example, a team could have a high win rate because they were lucky and got paired with good teams. Conversely, a team could have a low win rate because they were unlucky and got paired with bad teams. Therefore, win rate is not a good metric for measuring a team's performance.

Another metric that could have been used was average alliance score. Average alliance score is a better metric than win rate because it is a better indicator of a team's skill. For example, a team with a high average alliance score is likely a good team because they were able to score a lot of points. Conversely, a team with a low average alliance score is likely a bad team because they were not able to score a lot of points. Therefore, average alliance score is a better metric for measuring a team's performance.

However, average alliance score is not a perfect metric for measuring a team's performance. Average alliance score is flawed because it could still be affected by luck. For example, a team could have a high average alliance score because they were lucky and got paired with good teams. Conversely, a team could have a low average alliance score because they were unlucky and got paired with bad teams. Therefore, average alliance score is not a perfect metric for measuring a team's performance.

The best metric in my opinion would be something based off of contributions, such as average points contributed. This is because contributions are a better indicator of a team's skill than win rate or average alliance score. For example, a team with a high average points contributed is likely a good team because they were able to contribute a lot of points. Conversely, a team with a low average points contributed is likely a bad team because they were not able to contribute a lot of points. Also, a metric based on individual contributions is less likely to be swayed by luck. Good or bad alliances would likely have no effect on such a metric. Therefore, average points contributed is a better metric for measuring a team's performance than win rate or average alliance score.

A metric for a team's average contribution of some value x could be calculated as follows:

$$\text{contrib}(n) = \bar{x}_n - \bar{\bar{x}} + \frac{1}{3}\bar{\bar{x}}$$

where n is the team number, $\bar{x}_n$ is the average value of x for team n, and $\bar{\bar{x}}$ is the average value of x for all teams. This metric is calculated by subtracting the average value of x for all teams from the average value of x for team n and adding back $\frac{1}{3}\bar{\bar{x}}$. The extra $\frac{1}{3}\bar{\bar{x}}$ is added back with the assumptions that the grand mean x is contributed equally by all 3 teams in the alliance. It can be rewritten as follows:

$$\text{contrib}(n) = \bar{x}_n - \frac{2\bar{\bar{x}}}{3}$$

This metric is better than average alliance score because it is immune to lucky alliances.

7 Appendix

8 Bibliography