Project 1 Analysis

# **Formula 1 Racing Analysis**

1. **Introduction**

Formula 1 is the premium motor racing sport in the world and for years data has played a major role for teams and drivers looking to improve their times and results. It is widely believed that every second of the race matters which factors heavily into the race result for a driver. Drivers typically have to pit two to three times a race to change tires or make any other repairs to their car, these stops all cost the drivers precious seconds and we wanted to see if this is truly a main factor of a driver’s race result or if there are many other factors that contribute more strongly to their result. The other factor we are looking at is if the driver’s starting position affects their race result. Every Saturday of a Formula 1 race weekend, each driver competes for the fastest lap to see where they will start the race on Sunday. Again, it is widely believed that starting first in a race is incredibly important for winning, but we wanted to see if there was a correlation between the average starting position and their average finishing position. The final piece we wanted to investigate was races affected by the weather. With rainy races, the cars must drive slower, which evens out the grid. During these races, we wanted to see if the driver with the fastest pit stops had the best race results. We felt as if this would give us the best idea if faster pit stops were always better for drivers and truly affects their races because the car speeds are far more even in these rain-affected races.

1. **Question 1**
   1. How much does the average pit stop time of a driver per race affect their finishing position?
      1. Our null hypothesis for this question was that the average pit stop time of the driver for a race has no affect on their finishing position for that race. Our alternative hypothesis was that the faster the average pit stop time of the driver was for a race, the better their result will be for that race.
         1. For this question, we made two plots that show the average pit stop time for drivers vs their average finishing position throughout a season. For these plots, we excluded data that included the pit stop times from red flag sessions which is when the race is stopped for an extended period. These results were fascinating because there is somewhat of a positive correlation between faster average pit stop times and average finishing positions. For our scatter plot showing this information, the line of best fit had an r-value of 0.1966. This means there is a weaker relationship between the two variables, but it is still noticeable. However, there were a few drivers who had quicker pit stops, but not the best race results. Because of these results from our plots, we cannot make an accurate conclusion about either of our hypotheses. Pit stop times likely play a major role for drivers and their finishing position, but it is also more likely that other factors like strategy, aerodynamics, and car performance play just as large of a role in a driver’s race. For future research, we would want to look at more data throughout the seasons, but it is important to note that pit stop crews can change from season to season, so it would be best to compare season by season to make a conclusion.
2. **Question 2**
   1. How much does the starting grid position of a driver affect their finishing position?
      1. Our null hypothesis for this question was that average grid position for the driver has no affect on their race results in a season. Conversely, our alternative hypothesis was that the better average starting position a driver had in their season, the better race results they had in that season.
         1. Upon looking at our data and creating various data frames to compare these two factors, we found conclusive results. We began by grouping the drivers for the 2021 season and getting their average starting position and then their average finishing position. From there we created a scatter plot to display this data and plotted a line of best fit. The r-value we obtained from this plot was 0.8522, meaning there was a strong correlation between these two variables. This was clear from the scatter plot too because much of the data stayed near the line of best-fit meaning that the better position they started in the race, the better they finished in races. This trend remained consistent across all 20 positions. The next plot we created was similar to the last one but showed these two variables grouped together by constructors. We wanted to see this data because this shows us data for two of the same cars, but with different drivers in each car. This gives us a better picture of how car performance is crucial for these two factors. Once again, there was a strong positive correlation here as our r-value for this scatter plot was 0.9319. Again, much of the data was consistent through all the positions and the points of data remained near the line of best fit. For this question, we rejected our null hypothesis and accepted our alternative hypothesis. We can conclude that starting a race in a better position will lead to better-finishing results in that race.
3. **Question 3**
   1. How much does weather affect the average pit stop time and average driver finishing position?
      1. Our null hypothesis for this question was that weather has no affect on the race result and the drivers with the best car finish better in this race. Our alternative hypothesis was that the driver with the fastest average pit stop times in this race finished better than the driver’s who had slower pit stop times.
         1. Reviewing the data to this question did not provide much of a clear answer. We began by taking data from the 2021 Emilio Romagna Grand Prix and the 2021 Hungarian Grand Prix. We picked these two races because we believed the results were heavily influenced by the weather on the day of the race, which was a very rainy race. We made two scatter plots, one with the average pit stop time vs the average finishing position for these two races, and another with the average number of pit stops vs the average finishing position. The correlation between these variables across the two plots was weak to nonexistent and it is hard to make a conclusive decision. We think it is fair to say that weather-affected races are incredibly unpredictable, and any driver can win. The factors that influence a winner of a race may be more closely tied to the strategy of the driver, and the overall ability of the driver in the rain. Because of the results we saw in our scatter plots, we can accept our null hypothesis and reject our alternative hypothesis for this question.