

All About Pit Stops: An Analysis of The Japanese Grand Prix in Formula One

Isaac Wetmore - 1005756674

April 19, 2021

Abstract

The following report serves as an analysis of pit-stop times in the Formula One world championship for a the specific track Suzuka used for the Japanese grand-prix. It is often highly debated whether it is the drivers' skill, or their team's effort, that determines the success of that driver. This report hopes to break down how a team's pit crew and strategy choices can affect the performance of a driver. The analysis answers this questions in two parts. The first looks at average pit-stop times and attempts to develop a strong estimator for its mean, followed by a Bayesian analysis that highlights the credible interval of that mean based off the sample size. We then use linear regression to decode the relationship between average pit-stop time and performance, observing a moderate correlation. The second part of the analysis looks at whether one-stop or two-stop strategies are used more through a goodness of fit test, and whether one strategy is better than the other using a bootstrapped confidence interval. Our calculations showed that although both strategies are possibly used at equal intervals, a one-stop strategy is equally as effective or better than a two-stop strategy. We then discuss the implication of a one-stop strategy being better, and the correlation between pit-stop times and performance to conclude a statistically analysis of pit-stop times in Suzuka.

Introduction

Formula 1 is the highest competitive class in the single-seater category for competitive racing. It is enjoyed globally by millions as one of the most prestigious auto-racing competitions in the world. Since its debut in 1950, hundreds of drivers have competed in the sports history, funded by historic and legendary car manufacturers such as Ferrari, Mercedes, Honda, McLaren, BWM, Porsche, Renault, and Lotus. The current setup is 20 drivers on the grid, with 10 "constructors" (2 drivers per team) that handles development of the car and race strategy. The sport is famous not only for the skill required by the drivers, but also the efficiency of the teams to develop a car and pit crew that can maintain technically consistency over the course of over 20 races per season.

One very important aspect of Formula One is a team's pitting strategy. F1 currently permits the use of three different types of tires that can be used throughout the race: soft, medium and hard tires. The softest tire yields the fastest race times, but degrades quicker, while the hardest tire is the slowest but has the longest race life before requiring a "pit stop" to change. A large enjoyment from F1 comes from seeing each team set out a specific strategy they think will win the race. One strategy is keeping to medium and hard tires and stopping less at the cost of less pace, while others will stay on soft tires, pushing their cars to the maximum but requiring more stops. By rule, a driver must stop at least once (they must use two different types of tires during the race). Regardless of the rule, even one set of hard tires is not designed to make it much more than half the distance of any Grand Prix. Generally, a one-stop strategy is considered the most viable for many tracks, though there are exceptions. For the Japanese Grand Prix in Suzuka, a two-stop strategy is also considered viable by many analysts.

Not only is the strategy determined by the teams, but also the speed of the pit stop. On average, a standing pit-stop in Formula One is 2 to 3 seconds, with some of the fastest pit-stops going under 2 seconds (DHL Fastest Pit Stop Award, 2019). It is often claimed that a driver's race can be made or broken by how fast a constructors pit crew can accommodate them. This report will provide statistically evidence either supporting the claim that pit stops are very critical to a race, or suggest their importance is only trivial in the grand scheme of performance.

This report will analyze the pit stop times and strategy for the Japanese Grand Prix in Formula One. Keeping our focus to one track is made for several reasons. The most pressing reason is due to how pit stop times are tracked. A pit stop time begins when a driver has entered the pit lane, not when they have stopped for the actual tire change; therefore, pit stop times vary per circuit, all with different pitlane lengths. Moreover, Suzuka gives us a good overview of some of the best Formula One has to offer in terms of race strategy. Many tracks (including some of the most famous such as Monaco or Monza) have fairly straightforward one stop strategies with little variation. The variance in strategies present in the Japanese Grand Prix make it an excellent track to study.

The report will consist of five different methods that will each be discussed with more detail in the methods section. The first three will observe the distribution of average pit-stop times. The MLE estimation will derive a mean estimator for the data. A Hypothesis Test will highlight why metrics during broadcasting that attempt to predict pit stop times are sometimes inaccurate. The Bayesian interval test will showcase a possible range of values the average pit stop time for the track based on the sample size of 5 races we will use. This part of the report will conclude with linear regression that will attempt to illustrate any possible correlations between average pit-stop time and finishing position.

The second part of this analysis will analyze the details of choosing a one-stop strategy or a two-stop strategy. The goodness of fit test will be applied to determine whether the sample suggests teams choose a one-stop or two-stop at this track equally or at different rates. The report will follow with a bootstrap confidence interval that will showcase the average performance of both teams that use a single-stop strategy compared to those who use a two stop strategy. Assuming no major confounding variables, these results could showcase whether one stop is statistically better than the other, or if both yield equally strong results on average.

This report thereby contains two hypothesis. The first suggests that average pit-stop times will have an effect on performance, with even marginally changes in pit-stop times effecting where the racer will end up. The second hypothesis relates to whether a one-stop or a two-stop strategy is better. It is often stated that at Suzuka, a two stop strategy is better, therefore, my hypothesis is that we will see the two stop strategy picked more and it will yield better results on average.

Data

The data set comes from The Ergast Developer API, an web service that has compiled historical data for many motor sport events. The particular data set on F1 contains historical results dating back all the way to 1950. Although Ergast was the original source, the data-set I acquired comes from Kraggle (Vopani, 2021).

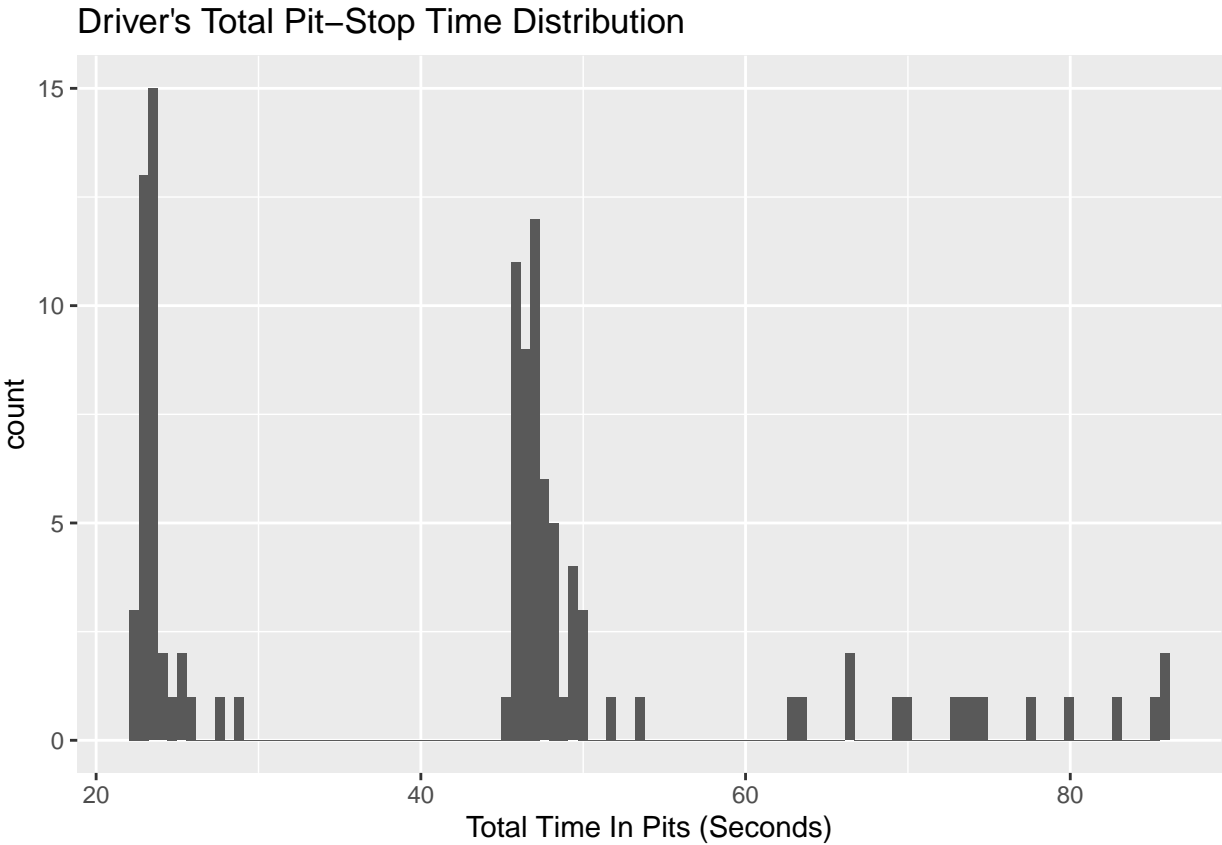
The data set was filtered to focus only on races at Suzuka from 2013 to 2019 with 2014 removed (5 races). Although Suzuka entered the calendar in the 80s, this report will only look at pit stop data occurring after 2012. This restriction is due to refueling during the race being banned in 2010, significantly changing the length of a pit stop. Before 2012, the pit lane speed limit was also faster, making the data inconsistent with the rest of the races. 2014 was removed as this was a race in the rain. Races in the rain cannot be used because generally drivers will pit more frequently to match the often quickly changing conditions on track, as such, it will only disrupt our analysis of whether a one or two stop strategy is the fastest. I also removed any driver that did not finish the race, as this would be another confounding variable.

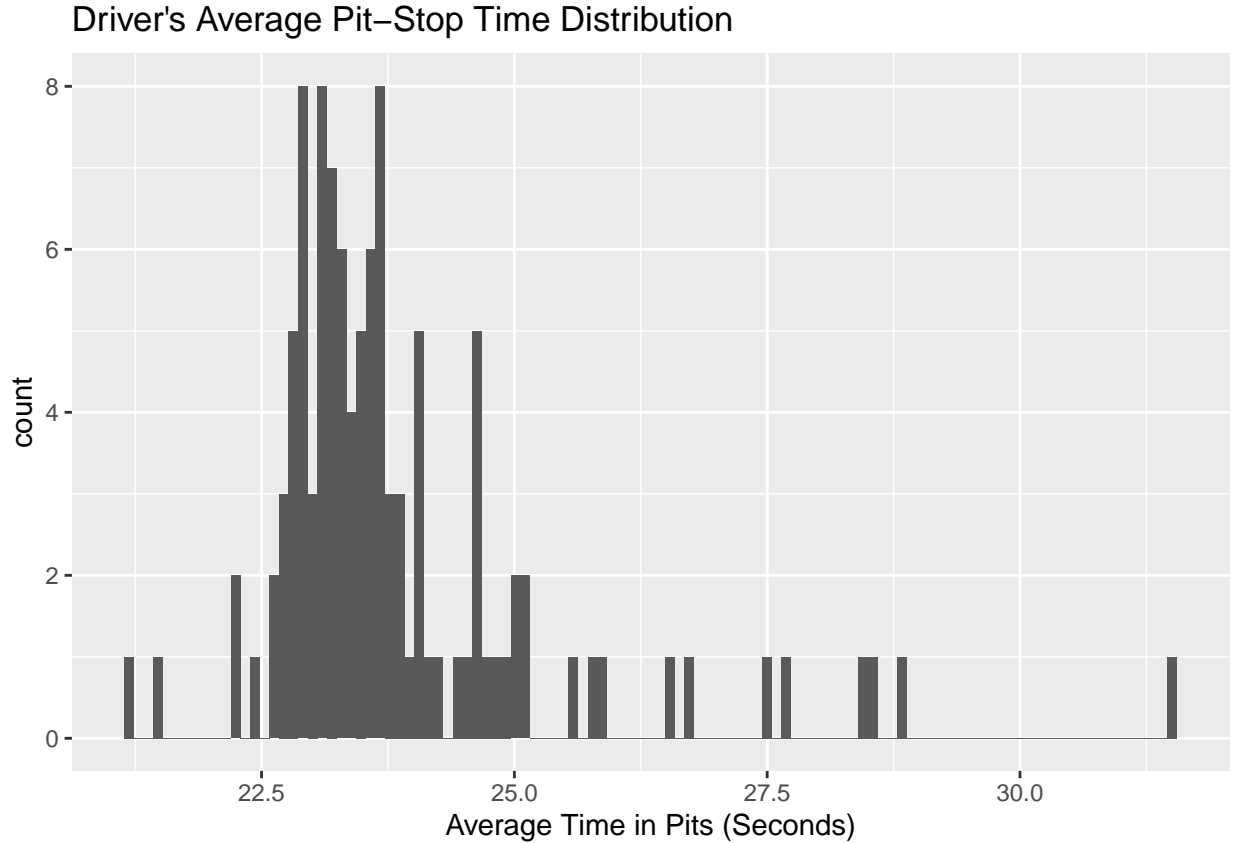
Two statistics have been derived from the data and added to the data set separately from the original source. The first has taken the average of the pit stops from each driver, and the second has summed the total time each driver spent in the pits. The average is more useful with analysis related to the performance of the pit crew on average, while the later is more useful for observing the effect of pit strategy.

There are four important variables to this analysis. The first two are average pit stop time and total pit stop time as discussed above. The fourth is how many stops the driver took during the race. The fourth is the finishing position of the driver. Our linear regression model will observe the relationship between the pit stop times and the ending position (1st, 2nd, 3rd, etc.).

Summary Statistics	
Mean for Average Pit-stop time (Seconds)	23.88
Standard Deviation for Average Pit-stop time (Seconds)	1.51
Mean for Total Pit-stop time	42.88
Standard Deviation for Total Pit-stop time	17.36
Number of Drivers Using One-Stop	39
Number of Drivers Using Two-Stop	55
Number of Drivers Using Three-Stop	14

Above is a summary table that showcases the means and standard deviations for average pit stop time for every driver. The average pit stop time at this track is 23.88 with a relatively small standard deviation. Also included is the distribution of the amount of stops each driver took, with more preferring a two-stop strategy, though a one stop remains close behind.





The two graphs above plots the count of total pit stop times and average pit stop times and respectively. For total pit-stop times, we observe a clear bimodal trend. This distribution makes sense; drivers who only commit to one-stops have their own normal distribution, followed by drivers with a who took two stops. Drivers who took three stops show no normal distribution. This result is likely because drivers who have to take three stops only do so if something went very wrong out on track, such as their tire puncturing or front wing being damaged from contact.

For the average time spent on each separate pit stop, we observe a clear normal-like distribution that is slightly skewed to the right. This suggests that for the sample set, the bigger factor in slower pit-stop times is messing up and taking more time then needed. There are very few stops that are faster than the peak (suggesting the mode is bigger than the mean).

Regardless, this analysis will precede under the assumption the data is normal. This assumption may negatively affect our analysis in multiple ways, but also makes for a more straightforward report. Linear regression will be affected the most, for the model is rested on a normal distribution. For both the MLE and Prior, the assumption this data as a normal distribution is less of a problem. The use of a normal in estimating an MLE will yield a reasonable results, and the posterior function derived from a normal prior will also yield useful intervals. We will discuss the weight of this assumption at each section.

Methods

The following is a more detailed outline of the different methodologies used in this report and their underlining assumptions.

Maximum Likelihood Estimator

We will assume the data is a random sample of Normal random variables with mean, μ and variance $\sigma^2 = 2.25$. I have used the maximum likelihood estimator (MLE) approach to estimate the mean, μ . The MLE of μ is \bar{X} . All derivations regarding the MLE can be found in Section 1 of the Appendix.

This assumption is fairly bold considering the data is likely skewed moderately to the right. However, the analysis will not be drastically off reality with this assumption considering the data is only mildly skewed at most. The estimator statistic for mean is still relevant despite the data being skewed. This flaw can be seen less as a flaw for the current analysis, and more a potential barrier to deeper analysis of other estimators within the data set.

The potential limits of this assumption will be further discussed as they come up during the analysis and in the conclusion. Note here that for the MLE deviation of the mean, this assumption is the least consequential of the three.

Hypothesis Test

During the broadcast of the race, it will often be mentioned that the average pit stop is on average 20 seconds. AWS will seemingly use this metric to predict statistics for how likely a driver is with overtaking another driver when they come in to the pits. How close is this metric to the actual average pit-stop time given in this sample however?

Observing some of the outlining data, we notice a few values of cars only taking 18 seconds through the pit. This result is likely due to “drive-through” penalties drivers will sometimes get during a race as punishment for some illegal driver action. Typically this penalty—though rarely given out—requires the driver to drive through the entire pit lane, not having to stop at their box. An interesting point revealed by these outliers however is it gives us an idea of how long it takes to drive through the pit lane without a stop—around 18 seconds. For the sake giving this metric a slightly more accurate estimate, we will say that this average is likely 20 seconds if you account for slowing the car down to the stop. It should be noted that this is a generous estimate however.

It is fairly contested what the *average* standing pit stop time really is. For 2019, statistics from DHL directly yielded an average stopped time of 3.27 seconds (<https://f1bythenumbers.com/2019-f1-season-the-pit-crew/>). Applied to Suzuka, it is reasonable to compare the average pit stop time of the sample size with the likely ideal pit stop time: 23.27 seconds (21.27 with two seconds generously added on to account for having to slow the car down from the speed limit). Outliers on the other extreme (very long pit stops) obviously distort this metric, but it does make for an interesting hypothesis test. What is the chance the mean average pit stop time for many races in the future will average to this 23.27 second mark? Out null hypothesis is that the mean average pit stop time is 23.27 seconds ($H_0 : \mu = 23.27$), and the alternative hypothesis is that the mean average pit stop time is not 23.27 second ($H_A : \mu \neq 23.27$).

Since the sample size of n is fairly large, we will use standard normal distribution instead of a t distribution as our test stat.

Bayesian Credible Interval

Developing the findings of the Hypothesis test, we can develop an interval of values the average pit stop times can take on through a Bayesian posterior deviation. As noted above we will be using a normal distribution as a prior. This may prove to be a faulty assumption, and will be discussed further in the conclusion.

Another assumption we are making is holding the variance of the distribution as constant (held to the value of the sample). This assumption is the result of avoiding heavier algebra than is necessary for this analysis. While it would technically be more realistic to derive a posterior distribution that looks at both variance and mean, since we are only calculating a credible interval for mean, we can hold the variance as constant.

We are interested in finding a 95% credible interval of the parameter μ . Suppose our data is a random sample of Normal random variables with mean, μ and fixed variance, $\sigma^2 = 2.25$; and the prior distribution of μ is assumed to be Normal($\mu_0 = 0$, $\sigma_0^2 = 1000$) in hopes of yielding a neutral/non-informative prior. The posterior distribution of μ is Normal with mean $\frac{36000n\bar{x}}{8+36000n}$, and variance $\frac{9000}{9+4000n}$. Thus, we can use the first and 97.5th percentiles of this distribution to derive a range of values, in which μ has 95% probability of falling into. All derivations regarding the posterior distribution can be found in Section 2 of the Appendix.

Linear Regression

This section will form the heart of this analysis by crafting an analysis of the relationship between average pit stop times and the finishing position of the driver.

To observe the relationship between two variables, we can use a linear regression model. One variable is assigned as Y (the ending position) while the other is assigned X (the average pit stop time). A linear regression line will then show how much the y variable changes with a change in x. The intercept can be interpreted as the mean value of Y for a set value of X (when x is 0). If the slope is positive, this suggests a positive correlation. If the slope is negative, this suggests a negative correlation.

$$Y_i = \alpha + \beta X_i + U_i$$

Where α is the intercept of the regression line, β is the slope. Y is the dependent variable and X is the independent variable. U_i represents the bias modifier which accounts for statistically descrepancy

As mentioned above, our assumption of a normal distribution rears its ugly head in this part of the analysis. For a linear analysis to be viable, a core inference that must be met is a normal distribution of data. We cannot confirm this inference, and therefore are results must be taken with this assumption fully in mind. As an attempt to partially normalize the data, I decided to transform the variables by log. This transformation will not completely remove the problem of a skewed distribution, but it will dampen it. It could be argued that this dampening makes the skewed data so marginal that the linear analysis mostly applies. It will be noted here that this decision is slightly unnecessary to begin with; running the regression without the transformation yields a similar equation with similar conclusions to draw from.

Goodness of Fit Test

An interesting lead in towards the rest of the results can involve investigating what pit stop strategies choose, regardless of its effectiveness. We will observe the categorical distribution of those who chosen one stops compared with those who chose 2-stops, and computer a Goodness of fit test to determine whether, for the population, these strategies are chosen equally or one is chosen over the other. If our null hypothesis is not rejected (H_0 : Both one stop and two stop strategies are chosen equally ($1/2, 1/2$)) then we can take the hypothesis as valid. If we do reject the hypothesis (H_A : Strategies are not chosen an equal amount), then we can conclude that a two-stop strategy is likely more popular. Note that for this calculation, I dropped three stop strategies as these are not considered viable and only done out of necessity from some problem—as is evident by so little drivers choosing them.

We will use a Chi-squared test to compute the uniformity of the distribution of pit stop strategies. This involves the test strat: $-2\log(\frac{L(p_0)}{L(\hat{p})}) \sim X_1^2$ With $p_0 = \frac{1}{2}$ and \hat{p} equally the distribution of the sample. the $L(p) = p_1^{n_1} * p_2^{n_2}$ where n is the sample's categorical representation in the sample and p is the probability of drawing that category (SOURCE).

Confidence Interval

This section shall complete the anaylsis by observing the difference in performance drivers who do one stop have with drivers who complete two stops. We will computer a confidence interval bootstrap that will sample

the results of drivers who had one stop with and subtract the mean of that sample with the mean result of drivers who took a two stop strategy. The interval of values will give us an idea of which strategy is statistically better.

Results

Maximum Likelihood Estimator

MLE Statistics	
Mean for Average Pit-stop time (Seconds)	23.88
Mean for Total Pit-stop time	42.88

The table above showcases the estimator used on the data set, yielding the same results as the `mean` function used for the summary statistics. Under the knowledge that the `mean` and `sd` functions above are based on the assumption of normally distributed data, we can verify this MLE as correct.

Hypothesis Test

Hypothesis Test	
Test Stat	2.83
P-value	0.005

Due to the extremely low p value, though just on the threshold, we can reject this hypothesis. This result could be indicative of two things. The first is that my estimate of the pit lane taking 20 seconds to traverse is incorrect. The second could be that pit stops at Suzuka are harder than at other tracks, though the reasons for this discrepancy are unknown. Regardless, the test showcases how precise the credible interval likely is (something we will analyze now).

Bayesian Credible Interval

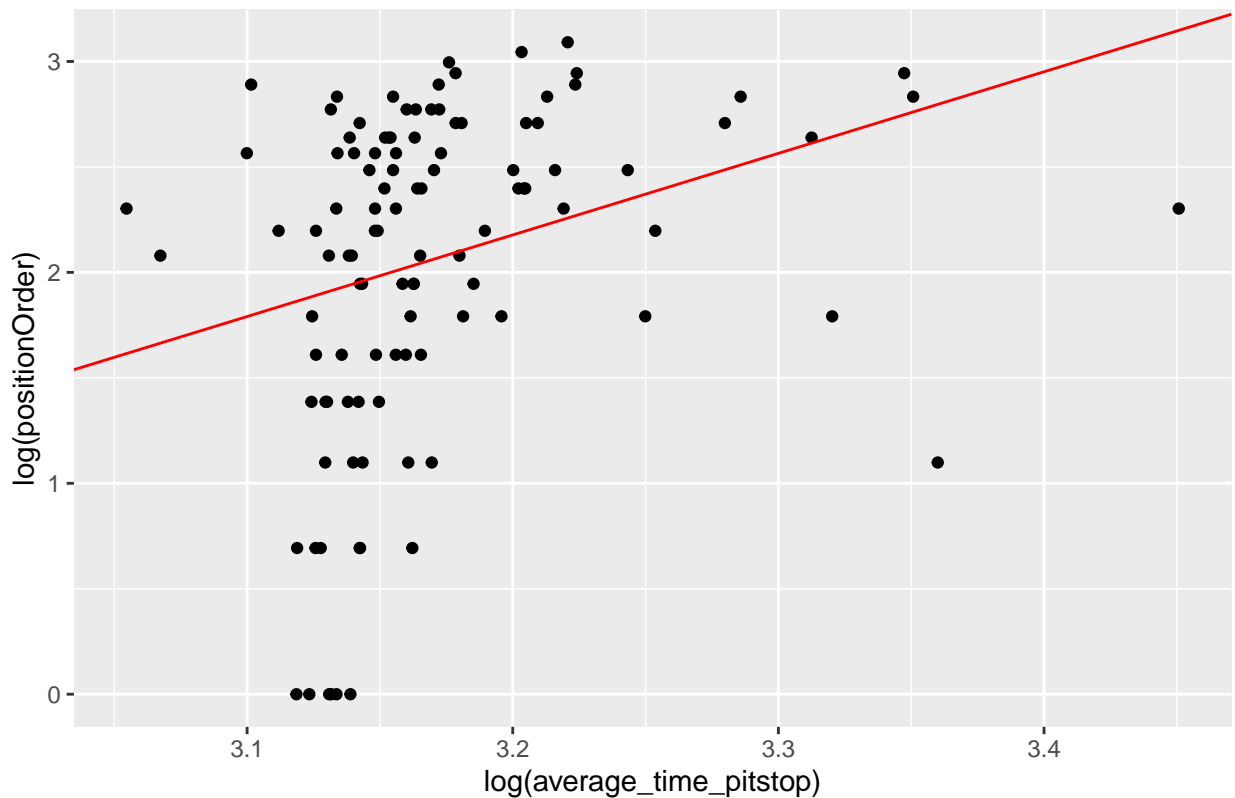
The calculated confidence interval of the resulting posterior distribution yields a confidence interval [23.6, 24.17] with 95% confidence. We can use this interval to make the reasonable estimate that the true average pit-stop time for the Suzuka track is in the range of 23.6 to 24.1 seconds with 95% accuracy. This fits in line with our previous hypothesis test, which found a value way off this range having a low chance of being the true interval.

Linear regression.

$\hat{\alpha}$	-10.2
$\hat{\beta}$	3.868

As one can note, there is a positive correlation between the length of pit stop and the ending position. It should be noted that position order was not reversed: the higher the position, the worse result. Although slightly confusing, the results essentially states that as the pit stop time increases by a certain amount, the ending position ends up being farther from 1st place. This result is likely not surprising, but what should be noted is the small margin of error present. A difference of even one second longer can influence one's order on the grid by 1-2 spots. That may not seem like a massive drop, but F1 gives out championship points at an exponentially lower rate with each drop in position; therefore, the difference between 5th place and 6th for the constructors championship is significant. The suggestion that a pit crew that is even one second too slow could be the cause of this drop is noteworthy.

Driver's Average Pit-Stop Time Vs Position Order (Log Transformation)



Above is a plot showcasing the linear regression model. Note that because the graph was transformed by log, the axis scale is not correctly fit to the actual data, but instead the log scale. It should be noted that the original regression model (without the log transition) is fairly similar to the one above; however, this plot is technically more accurate and was therefore chosen. As should be noted, we observe a clear predictive uptrend. Although the margins are small and there exists many exceptions, the general trends reveal to us a definite correlation.

Goodness of Fit Test

Goodness of Fit Test	
Test Stat	2.737
P-value	0.098

The P-value essentially communicates to us that there is a 9.8% chance the first and second pit stop is chosen equally. This result is not low enough to reject the null hypothesis. Although from the results a two stop seems slightly more likely, there is still a strong chance teams will equally choose between a one stop and two stop. These results will pair nicely with our confidence intervals below that outline the performance of these two different strategies.

Confidence Interval

The intervals computed come out to be $[-3.29, 0.52]$ showcasing that if we were to repeat the bootstrap, 95% of the confidence intervals derived would fall between these values. It is important to note how to interpret these values. The negative means that the two-stop strategy had a higher average position than the one-stop, which implies that group is finishing worse on average. A two stop strategy is correlated as being, at worst, 3 places *higher* (worse) than those who opt for a one stop strategy. At best, a two stop strategy can yield a half a place better on average.

Conclusions

The following report analyzing everything to do with Pit-stops at the Japanese Grand prix by analyzing statistics related to pit-stop times over the course of five races. We used a MLE deviation to find an estimator for the mean, which we used in calculating summary statistics and for other parts of the analysis. We then both a Bayesian posterior test and an hypothesis test to derive more information about the mean average pit stop time, which we found to be within a very tight range of about half a second. The chance of that average deviating from that 0.5 second interval was found to be exceedingly unlikely as shown by the hypothesis test. We then observed the relationship between the average pit stop time and finishing position, observing a mild to moderate correlation. Even a second slower on a driver's pit stop could affect their finishing position. These results highlight the importance in teams investing in a strong pit-crew and that even factors that seem like an afterthought can potentially have significant influence on a driver's success. There are likely teams that have strong pit-crews but weak cars, for example, but there are also likely teams that could be scoring better but have weaker pit-crews that are holding them back.

The second part of the analysis utilized a goodness of fit test and confidence intervals to estimate both how popular a one-stop strategy was compared to a two-stop, and the respective performance of both. We observed a mildly unlikely but reasonable chance of a one-stop and two-stop strategy being chosen equally, and observed that on average, but found that One-stop performs as well to or possibly even better than a two-stop strategy on average.

There are, however, other explanations for these results other than simply concluding that a one-stop strategy is likely superior. Teams who start father back are often incentivised to choose two-stop strategies over one-stop strategies.

Qualifying takes place the day before a race to determine the starting order for that race. It is set up as a three-round knockout (Q1, Q2, and Q3) with the first sorting the bottom 5 times, the second round sorting places 10-15, and the third round determining the order for the top 10 positions. For example a driver can get the 2nd fastest time in Q1, which allows them to move to Q2. That same driver can then produce the 13th fastest time in Q2, eliminating them from Q3 and locking in thier starting order for the race at 13th.

A key addition to strategy is that the tires used to set their Q2 times (soft, medium, or hard) will be that driver's starting set for the race. Generally, the slower teams will be fighting to even make it to Q3, opting for the soft tires. Starting the race on soft tires, however, basically ensures the driver *must* take a two stop strategy at Suzuka. The fastest drivers with better cars, in contrast, generally have much more breathing room. As such, they tend to use Medium tires in Q2 with little fear of being knocked out, allowing them to complete a three stop strategy.

This explanation may help explain why many choose two-stop strategies even if they yield worst results. The midfield teams are biased into choosing two-stop strategies, while the winning teams have much more freedom.

This explanation only can bring us so far, however, as if you do not make it to Q3, you can start on any tire you want. This essentially confirms that much of the bottom 10 still have a choice of mediums. So while this explanation may partially explain the results, a more advanced analysis of these variables is likely required to develop a better picture.

Weaknesses

As has been mentioned multiple times now, assuming average pit-stop times followed a normal distribution did limit the extensibility of the analysis in some ways. Other distributions that better accounted for the skewed data set would have been more appropriate. Moreover, even within this normal analysis, there were still a lot of assumptions made. The biggest one was holding the variance of the distribution as fixed while only looking at the mean. A more accurate derivation of the posterior distribution with an unknown variance would have involved even more algebra than was used. The prior parameters we chose were also fairly arbitrary and our results may have been influenced by different choices. Having a more accurate and flexible prior would have likely have changed our posterior distribution slightly and given us a much more specific confidence interval than what we got. It would have also allowed us to provide an interval on the variance of all race average pit-stop times instead of just the data values.

Moreover, our linear regression model was also flawed. Linear regression rests on the data being normal, with the distribution of even the data set showing signs of being skewed to the right. Since a normal distribution is defined by having equal sides, this deviation could have distorted the results of our regression. There are models superior to linear regression that could have been use to predict the correlation between pit-stop times and ending results.

For the sake of what we set out to achieve, however, the assumptions were not severely damaging to any conclusions we made in this analysis. Regardless if we had better accounted for the true distribution for the data set, the conclusions from our MLE derivation would have been identical, our Posterior deviation would have been close to identical, and our linear regression would have differed slightly but still would have supported the correlation.

Next Steps

Future work likely involves utilizing better prior distributions and regression models to derive more accurate predictive models. There are other distributions that exist that better account for this skewing. Furthermore, even if we wanted to keep a normal distribution, we could have set the variance as unknown to deepen the analysis.

In regards to the data itself, we highlighted above that there are other confounding variables that may affect the pit-stop strategy other than merely performance. Although this report provides a good foundation, future analysis would involve a redesign of some of the methodology to better account for these confounding variables. For the confidence interval, for example, a redesign of methodology could remove the Q2 bias to create a more interesting interval.

Discussion

Regardless of its shortcomings, I believe this report has done a good job at highlighting some the depth of analysis that is required to be successful in Formula One, teams work with leagues of data such as car technical data (telemetry data), lap-times, tire times, and more to craft predictive models of where their cars perform and where they need work. Although pit-stop data is admittedly a smaller aspect of F1, the depth of analysis we were able to produce from it serves as a testament to the complexity of the sport. If this report got you even slightly intrigued by the strategy and technical preciseness and sophistication of F1, or any motor sport such as Nascar or Indycar, I consider this report successful.

Bibliography

1. Golemund, G. (2014, July 16) *Introduction to R Markdown*. RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: January 15, 2021)
2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.
3. Allaire, J.J., et. el. *References: Introduction to R Markdown*. RStudio. <https://rmarkdown.rstudio.com/docs/>. (Last Accessed: January 15, 2021)
4. Evans, M., & Rosenthal, J. S. (2010). Chapter 7 - Bayesian Inference. In *Probability and statistics: the science of uncertainty* (pp. 373–431). essay, W.H. Freeman and Co.
5. (2019, December 10). 2019 F1 season: The pit crew. F1 by the Numbers. <https://flbythenumbers.com/2019-f1-season-the-pit-crew/>.
6. DHL Fastest Pit Stop Award. DHL InMotion. (n.d.). <https://inmotion.dhl/en/formula-1/fastest-pit-stop-award-2019/>.
7. Vopani. (2021, April 6). *Formula 1 World Championship (1950 - 2021)*. Kaggle. <https://www.kaggle.com/rohanrao/formula-1-world-championship-1950-2020>.
8. Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
9. Joseph Larmarange (2020). labelled: Manipulating Labelled Data. R package version 2.7.0. <https://CRAN.R-project.org/package=labelled>

Appendix

Section 1

We know that for a likelihood function for a normal distribution (taken from lecture):

$$L(\sigma^2|x_1, \dots x_n) = f_{X_1, \dots X_n}(x_1, \dots x_n)$$

$$L(\sigma^2, \mu) = \prod_{i=1}^n \left(\frac{e^{-1/2(\frac{x_i - \mu}{\sigma})^2}}{\sqrt{2\pi\sigma^2}} \right) \quad (1)$$

$$L(\sigma^2, \mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-1/2(\frac{x_1 - \mu}{\sigma})^2} * e^{-1/2(\frac{x_2 - \mu}{\sigma})^2} * \dots e^{-1/2(\frac{x_n - \mu}{\sigma})^2} \quad (2)$$

$$L(\sigma^2, \mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2}((x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 \dots)} \quad (3)$$

$$L(\sigma^2, \mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2} \quad (4)$$

The loglikelihood is the natural logarithm.

$$\ln(L(\sigma^2, \mu)) = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2}\right) \quad (5)$$

$$\ln(L(\sigma^2, \mu)) = n * \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \quad (6)$$

$$\ln(L(\sigma^2, \mu)) = -n(\ln(\sqrt{\sigma^2}) + \ln(\sqrt{2\pi})) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \quad (7)$$

$$\ln(L(\sigma^2, \mu)) = -n\left(\frac{1}{2}\ln(\sqrt{\sigma^2}) + \frac{1}{2}\ln(2) + \frac{1}{2}\ln\pi\right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \quad (8)$$

Since we know σ^2 is 2.25, we can substitute it in:

$$\ln(L(2.25, \mu)) = -n\left(\frac{1}{2}\ln(\sqrt{2.25}) + \frac{1}{2}\ln(2) + \frac{1}{2}\ln\pi\right) - \frac{1}{3} \sum_{i=1}^n (X_i - \mu)^2$$

To find the MLE, we must find the derivative of either the likelihood or loglikelihood:

$$0 = \frac{d}{d\mu} \left(-n \left(\frac{1}{2} \ln(\sqrt{2.25}) + \frac{1}{2} \ln(2) + \frac{1}{2} \ln \pi \right) - \frac{1}{3} \sum_{i=1}^n (X_i - \mu)^2 \right) \quad (9)$$

$$0 = \frac{2}{3} \sum_{i=1}^n (X_i - \mu) \quad (10)$$

$$0 = \sum_{i=1}^n (X_i - \mu) \quad (11)$$

$$0 = \sum_{i=1}^n (X_i) - n\mu \quad (12)$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad (13)$$

$$(14)$$

To prove this is the maximum, we can complete a second derivative test.

$$0 > \frac{d}{d\mu} \left(\frac{2}{3} \sum_{i=1}^n (X_i - \mu) \right) \quad (15)$$

$$0 > -\frac{2}{3} \quad (16)$$

Since the equation holds true, the above MLE is the maximum.

Section 2

Not that the proof involved in this section is greatly inspired by Chapter 7 of *Probability and statistics: the science of uncertainty* (Evans & Rosenthal, 2010).

The likelihood function derived from section one can be used:

$$L(2.25, \mu) = \left(\frac{1}{\sqrt{4.5\pi}} \right)^n e^{-\frac{2}{9} \sum_{i=1}^n (X_i - \mu)^2}$$

We will also need the prior distribution:

$$P(\mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}}$$

$$P(\mu|data) = \frac{P(data|\mu)P(\mu)}{P(data)} \quad (17)$$

$$\propto P(data|\mu)P(\mu) \quad (18)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} * \left(\frac{2}{\sqrt{9\pi}}\right)^n e^{-\frac{2}{9}\sum_{i=1}^n (X_i-\mu)^2} \quad (19)$$

$$\propto e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{2}{9}\sum_{i=1}^n (X_i-\mu)^2} \quad (20)$$

$$= e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{2n}{9}(\bar{X}-\mu)^2} \quad (21)$$

$$= e^{-\frac{1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2) - \frac{2n}{9}(\bar{x}^2 - 2\mu\bar{x} + \mu^2)} \quad (22)$$

$$= e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)\left(\mu^2 - 2\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{4n}{9}\bar{x}\right)\mu}{\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)}\right) - \frac{2\mu_0^2}{9} - \frac{n\bar{x}^2}{2\sigma_0^2}} \quad (23)$$

$$= e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)\left(\mu - \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{4n}{9}\bar{x}\right)\mu}{\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)}\right)^2 - \frac{1}{2}\left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{4n}{9}\bar{x}}{\frac{1}{\sigma_0^2} + \frac{4n}{9}}\right)^2} e^{\frac{2\mu_0^2}{9} - \frac{n\bar{x}^2}{2\sigma_0^2}} \quad (24)$$

$$\propto e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)\left(\mu - \frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{4n}{9}\bar{x}\right)\mu}{\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)}\right)^2} \quad (25)$$

$$(26)$$

The following setup lends us to a normal distribution with the parameters:

$$N\left(\left(\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{4n}{9}\bar{x}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)}\right), \left(\frac{1}{\sigma_0^2} + \frac{4n}{9}\right)^{-1}\right)$$

For the analysis in the Bayesian section, we chose the prior parameters to be $\mu = 0$, $\sigma_0 = 1000$ as it yields the best chance of providing a non-informative prior. A smaller variance would produce a wider credible interval, as the variance of the distribution would be larger. A larger mean would have slightly skewed our data (the more normalized the better), so 0 is the best choice.